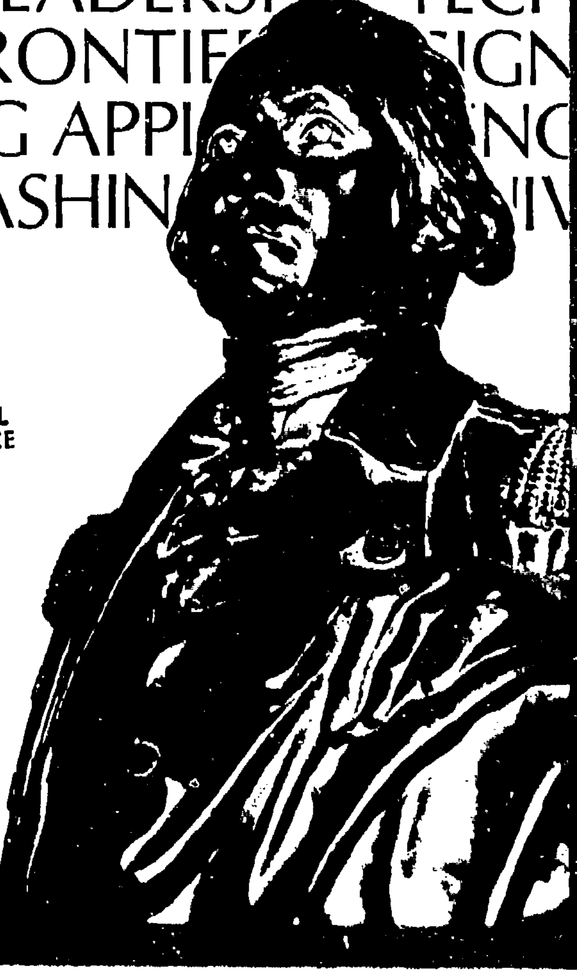
THE
GEORGE
WASHINGTON
UNIVERSITY

STUDENTS FACULTY STUDY R
ESEARCH DEVELOPMENT FUT
URE CAREER CREATIVITY CO
MMUNITY LEADERSHIP TECH
NOLOGY FRONTIE IGN
ENGINEERING APPL NC
GEORGE WASHIN IV



INSTITUTE FOR MANAGEMENT
SCIENCE AND ENGINEERING
SCHOOL OF ENGINEERING
AND APPLIED SCIENCE

EFFECTIVE STATISTICAL TESTS FOR DETECTION MODELS

by

James A. Lechner

Serial T-258
17 February 1972

The George Washington University
School of Engineering and Applied Science
Institute for Management Science and Engineering

Program in Logistics
Contract N00014-67-A-0214
Task 0001, Project NR 347 020
Office of Naval Research

This document has been approved for public
sale and release; its distribution is unlimited.

NONE
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* THE GEORGE WASHINGTON UNIVERSITY PROGRAM IN LOGISTICS	2a. REPORT SECURITY CLASSIFICATION NONE 2b. GROUP

3. REPORT TITLE

EFFECTIVE STATISTICAL TESTS FOR DETECTION MODELS

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
SCIENTIFIC

5. AUTHOR(S) *(First name, middle initial, last name)*

LECHNER, JAMES A.

6. REPORT DATE 17 February 1972	7a. TOTAL NO. OF PAGES 51	7b. NO. OF REFS 12

8a. CONTRACT OR GRANT NO. N00014-67-A-0214	9a. ORIGINATOR'S REPORT NUMBER(S) T-258
b. PROJECT NO. NR 347 020	
c.	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d.	

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Office of Naval Research

13. ABSTRACT

A "Detection Model" is an entity which calculates an instantaneous probability of detection of a "target" by a "hunter," from the values of variables which describe the environment and the actions of both hunter and target, including past history, if appropriate. Given such a model, and a succession of non-identical trials which terminate at detection or after a given period of time (whichever occurs first), it is desired to test the adequacy of the model.

An approach to this problem is presented, based upon recognizing the set of trials as a non-homogeneous Poisson process. Ways to improve the "power" of such tests by rearranging various segments of the trials are presented and discussed, including proper implementation of the tests using a digital computer. Extensions to the problem of improving the model and/or devising a new model are briefly discussed.

NONE
Security Classification

NONE
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Poisson Process						
Generalized Poisson Process						
Detection Models						
Detections						
Testing						
Statistical Tests						

NONE
Security Classification

THE GEORGE WASHINGTON UNIVERSITY
School of Engineering and Applied Science
Institute for Management Science and Engineering

Program in Logistics

Abstract
of
Serial T-258
17 February 1972

EFFECTIVE STATISTICAL TESTS FOR DETECTION MODELS

by

James A. Lechner

A "Detection Model" is an entity which calculates an instantaneous probability of detection of a "target" by a "hunter," from the values of variables which describe the environment and the actions of both hunter and target, including past history, if appropriate. Given such a model, and a succession of non-identical trials which terminate at detection or after a given period of time (whichever occurs first), it is desired to test the adequacy of the model.

An approach to this problem is presented, based upon recognizing the set of trials as a non-homogeneous Poisson process. Ways to improve the "power" of such tests by rearranging various segments of the trials are presented and discussed, including proper implementation of the tests using a digital computer. Extensions to the problem of improving the model and/or devising a new model are briefly discussed.

THE GEORGE WASHINGTON UNIVERSITY
School of Engineering and Applied Science
Institute for Management Science and Engineering

Program in Logistics

EFFECTIVE STATISTICAL TESTS FOR DETECTION MODELS

by

James A. Lechner

Introduction

This report presents a new technique for analyzing the results of operational tests of complicated equipment or systems, considerably more efficiently than other techniques in the literature. It is applicable whenever the aim is to test a model of a stochastic system, where operation of the system affects the occurrence of certain random events, and the model purports to predict these effects. For example, the system could be a monitor-and-repair subsystem, with the events being failures of the overall system being monitored. Or the system could be a technique for deciding where to drill for oil, with the events being successful strikes. Finally, the system could be a detection system -- for aircraft, or ships, or burglars -- with the events being detections.

The model is treated herein as a "black box" -- i.e., no internal scrutiny of the model is performed. The evaluation is done purely on the basis of the model's predictions and the actual results achieved. It is assumed that the model operates on certain variables which describe the environment; it produces as output the "event rate" -- i.e., a function of time $\lambda(t)$ with the property that the probability of an event occurring in a small interval $(t-h,t)$ is $h \cdot \lambda(t)+$(terms of order less than h).

Previous test methods have been based on the idea of identical trials, for which the probability of a detection by time t is the same for each trial. When this conditio.. ;s satisfied, the times of detection for a set of trials can be considered as a sample from the cumulative distribution function (CDF) of time to detection, and the Kolmogorov-Smirnov test can be used. If not all trials result in detection, complications ensue -- but these can be handled, at the expense of having a conservative test with less power. The real problem occurs when the test conditions are not identical. In previous work, e.g., Reference [2], this was handled by taking the most conservative approach possible: all possible extensions of the test runs were considered, and the model rejected only if every extension resulted in rejection. Needless to say, this makes rejection unlikely -- whether or not the model is correct. Thus the test is both conservative (with respect to "size", which is the chance of deciding that the model is incorrect when it is correct) and less powerful than it could be. Also, this procedure does not allow any special attention to be paid to the effect of other variables, either in or out of the model. Both these shortcomings are overcome in the method presented in this report.

The method to be presented has two forms. The simpler form involves a transformation of time, to make the outcome of the trials look like a Poisson process (under the hypothetical condition that the model is entirely correct). The more powerful form involves an additional step, that of considering the trials in small segments of time and rearranging those segments to achieve greater power -- i.e., a greater chance of deciding that the model is incorrect when it is incorrect -- without changing the size. It may seem strange that one would consider breaking up and rearranging trials, especially if some of the variables determining the event rate depend on earlier time -- as for instance when the probability of detection depends on how long the signal strength has been above a certain threshold. There is no problem, however. The method applies in a straightforward way. In fact, as explained in the report, one may well wish to use those "memory" variables along with the instantaneous ones in deciding how to rearrange the segments. Even the simpler form of the method is superior to the former methods, in that it is not necessary to look at all possible extensions of the terminated trials. The more powerful form is a (considerable) further improvement.

The remainder of this report is arranged in several sections, as follows. Section 1 describes the entity known as a "detection model". (The applications to other kinds of problems, as mentioned above, require changes in wording.) Section 2 contains a short summary of the theory and practice of statistical testing, to ensure that the vocabulary of the rest of the report means the same thing to the reader as to the writer. Section 3 applies the statistical theory to the Poisson process (defined therein), and describes the advantages of rearranging segments. Section 4 shows how the general case of a detection model and given trials can be reduced (by a transformation of time) into the Poisson case, thus making the results of Section 3 applicable to the real problem. Section 5 considers the practical question of how to rearrange the segments, and also the details of carrying out the test once the data are at hand -- including the efficient implementation of the method on an electronic computer. Section 6 is a short discussion of the possibility of using the method in a diagnostic mode, to identify weaknesses of the model, and in a model-building mode, wherein regression techniques are used to produce a better model. Finally, in Section 7, the method is applied to sample data to exemplify the points covered in the earlier sections.

1. Detection models

The physical problem. There is a "target" (or targets), and a "hunter." Trials are performed, in which the hunter attempts to detect the target. The probability of detection is not necessarily unity for each trial, so that some trials may not achieve detection. These trials are not instantaneous events, but rather last for a period of time; detection could conceivably occur at any time during a trial. Occurrence of a detection terminates the trial. Records are kept of conditions obtaining during the trials, including times of detection, actions taken by hunter and target (if any) and environmental conditions germane to the trials. These data are to be analyzed to test a "detection model" which purports to give detection probability as a function of environmental conditions and the behavior of hunter and target.

The detection model. An entity known as a "detection model" exists, which accepts as input the values of certain specified variables (continuous and/or discrete, including categorical) and delivers a "probability of detection" value. The model purports to represent the physical situation, as it affects the probability of detection. The detection model is treated as a deterministic "black box" in this report -- i.e., for given input values it will always produce the same output values -- and it is not subject to internal scrutiny. It is, however, subject to test: its predictions are to be compared with actual trial results, with the intention of deciding whether or not the model does, in fact, constitute an adequate predictor.

Restrictions applicable to this report. The methods discussed herein are applicable to detection models satisfying one condition: the model specifies a "detection rate" $p(t)$, such that the chance of detection in a given small increment of time $(t, t + dt)$ is $p(t)dt$ plus terms of smaller order in dt . Nothing need be assumed about independence from one instant to the next, and the function p(t) may depend in any manner on the past and/or present. The tests are valid (in the strict statistical sense) with no more conditions. It will be pointed out below that, in contrast to the validity, the effectiveness of the tests depends on our knowing and using more about the model, and, in particular, about the ways in which the model might be incorrect, but this is inescapable, since the "power" of a test is necessarily a function of the true state of affairs. To put it more descriptively, it is easier to find something if you know what it is you are looking for. (It should be stressed that this truism applies equally to any statistical test procedure.)

2. Statistical tests of hypotheses: a summary

Statistical tests, as developed by R. A. Fisher, involve the following elements: (1) a null hypothesis, H_0, which is to be tested; (2) a specified protection level, α, also known as the significance level or the size of the test; (3) a statistic Y (in the sense of a function of data, the value of which will be completely determined once the data are collected), and a rejection region or critical region R chosen so that if H_0 is true, $Pr(Y \epsilon R) \leq \alpha$. In words, R is chosen so that under H_0, there is at most a small chance α that Y will take on a value in the critical region R. The procedure is then to take the data, calculate Y, and reject H_0 if (and only if) $Y \epsilon R$. Naturally, one tries to choose Y and R so that when H_0 is not true, Y will end up in R with high probability. Note that there are two possible types of error: rejecting H_0 when it is true, and not rejecting H_0 when it is false. The first of these is called the Type I error, and is limited by the specified value of α. The second is (logically enough) called the Type II error, and is often a much more nebulous entity. Actually, one more often sees reference to the power of a test, defined as one minus the probability of a Type II error. Thus α is the probability of rejecting H_0

when true, and the power (henceforth denoted by β) is the probability of rejecting H_0 when false. If under H_0 the distribution of Y is completely specified, H_0 is said to be a simple hypothesis; otherwise, H_0 is a composite hypothesis. If H_0 is simple, then the probability of a Type I error is a single well-defined number. Since the alternative to H_0 is not usually a simple hypothesis, the power is not usually a single number, but rather a function defined over all the possible alternatives to H_0 .

There are various refinements and extensions of the barebones outline just given. In some cases, uniformly most po''erful (ump) tests can be found. As the name implies, such a test is at least as powerful as any other test (of the same size) for each possible alternative under consideration. There are likelihood-ratio tests, and various statistics Y , as well as ways to obtain the sample sequentially so as to minimize the maximum expected sample size, or to minimize the expected sample size under H_0 and a particular alternative H_1 simultaneously, among all tests meeting both α and a specified power at H_1 . For our present purposes, however, the data collection plan is assumed given, so that sequential methods are of no concern, and the structure of the set of possible alternatives to H_0 is not specified well enough to make use of the other refinements mentioned.

A trivial example will make clear the necessity of considering the power of a test. If we set α at (say) 5% , then a very simple test of proper size is as follows. Choose a two-digit random number; reject H_0 if and only if the chosen number turns out to be less than 05 . This test has $\alpha = .05$, of course; unfortunately, it also has $\beta = .05$ regardless of the true state.

Consider next the problem of testing the hypothesis that a given pair of dice is "fair." This example will show the importance of considering possible alternatives to the null hypothesis.

One could toss the dice many times and compare the average number of "spots" showing with the theoretical value of seven. This would be a good test against alternatives involving bias in favor of low numbers (or high numbers) of spots. However, such a test would have very low power against the alternative that there is a higher-than-normal probability of getting a seven, with all the other probabilities deflated accordingly. The reason is that the distribution of the test criterion (Y) is very much the same under the alternative as under H_0 : the average number of spots is seven in either case, and the variance is even less under the alternative than under H_0 . Obviously, one would choose a different criterion if this alternative were suspected; perhaps the sample frequency of seven spots.

Note that one might consider examining all the sample frequencies, one at a time. There are 11 such frequencies, not independent (since they add to unity). It is important to recognize that when several tests are done, the size (or significance level) must be interpreted cautiously: if ten tests were done at 5% each, and they were independent, then under H_0 the chance of finding one or more significant results is not .05 , but rather $1 - (.95)^{10} = .40$! To achieve 5% , one would have to do the individual tests at 0.51% , or alternatively, to reject H_0 only when at least two, or at least three, tests give significant results. (Under H_0 , two or more significant results will occur with probability 0.086 ; three or more, with probability 0.0115.) The former is preferred when it is expected that only one test will be an efficient detector for whatever might be wrong with the dice; the latter, when it is suspected that several will be affected. Of course, there are many other tests that could be used. This example is discussed here simply to point out the very important interrelation between alternatives suspected and tests chosen.

One other consideration should be mentioned. The most powerful test will not be used if it cannot be implemented. Similarly, a slightly less powerful test is to be preferred if it is so much simpler that it can be used with more data (because it will thus have more effective power).

3. Tests for a Poisson process

The Poisson process. A stochastic process is a random function of one or more variables, the simplest case being a function of one variable. If there is only one variable, and it is ordinary time, then we speak of a continuous-time process. A Poisson process is an integer-valued, continuous-time stochastic process $Y(t)$ such that for any set of t_i satisfying $t_1 < t_2 < \ldots < t_n$, the set of differences $\{Y(t_{i+1}) - Y(t_i)$, $i = 1$, 2 , ..., $n-1\}$ are independent and Poisson-distributed with means $\lambda(t_{i+1} - t_i)$ respectively. The constant λ is called the "occurrence rate" of the process. An alternative characterization is that given the function $Y(t)$ up to time t_0 , $\Pr(Y(t_0 + h) - Y(t_0) = 1) = \lambda h + o(h)$, where the symbol $o(h)$ means terms of order less than h (as $h \to 0$) . One example of a Poisson process is the number of events occurring by time t when the time between events has a negative exponential distribution, e.g., the number of counts registered on a Geiger counter under certain conditions, or the number of failures by time t in a system with exponentially-distributed lifetimes and instantaneous repair. Our immediate interest in this process stems from the fact that it is a special case of a detection model, especially easy to handle -- and that the general case can be transformed into this case!

In the remainder of this section, tests for a Poisson process are described. The following section is devoted to showing how the real model can be treated with the Poisson theory, with later sections devoted to practical applications, extensions, and an example. For a more detailed discussion of the Poisson process, see e.g., References [7], [10].

The simplest case: testing a single observation period. Suppose $Y(t)$, $0 \leq t \leq T$, is asserted to be a realization of a Poisson process with occurrence rate λ, and we would like to test this assertion. How do we proceed? One reasonable test statistic is just the number of occurrences in $(0,T)$, which should be distributed as a Poisson variable with mean $m = \lambda T$. If this hypothesis should be rejected, we might still want to entertain the less restrictive hypothesis that the process is Poisson with unknown occurrence rate. Then the number of occurrences no longer tells us anything, so we ask about the distribution of the occurrences across the interval $(0,T)$, given the number n of occurrences. And it turns out (Reference [7] or [10], for example) that the times t_1, t_2, ..., t_n are distributed like the order statistics from a sample of n from a *uniform* distribution on $(0,T)$; or equivalently, the values $s_i = t_i/T$ should look like an ordered sample from a uniform distribution on $(0,1)$. There are two well-known, respected tests of this hypothesis: the Kolmogorov-Smirnov test, henceforth called the K-S test, and the Cramér-von Mises test.* The statistic to be calculated for the K-S test is simply the maximum absolute difference between the sample cumulative distribution function (cdf) and its expected value. The Cramer-von Mises statistic is

$$W^2 = \frac{1}{12n^2} + \frac{1}{n} \sum_{i=1}^{n} \left\{ s_i - \frac{2i-1}{2n} \right\}^2 .$$

*There exist tests more powerful than the K-S test, but less well-known. One promising candidate was presented in a paper given by J. M. Finkelstein and R. E. Schafer of Hughes Aircraft Company at the annual meeting of the Institute of Mathematical Statistics, August 1971, Fort Collins, Colorado, entitled "Improved Goodness of Fit Tests."

Discussions of the use of both tests, with examples, can be found in References [4] and [8]. Until recently, only the asymptotic distribution of the Cramer-von Mises statistic has been tabulated (in References [3], [9]). Only the K-S statistic is considered throughout the rest of this report. It should be understood, however, that the K-S test is only one of many possible tests that could be used (References [3], [4], [5], [8], [12]); different tests suit different purposes, as one might expect. Since the sample cdf has jumps of size n^{-1} at the points s_i and is otherwise flat, and its expected value is simply $s(0 < s < 1)$, the maximum absolute difference can be expressed as $D = \max_i \left\{ s_i - \frac{i-1}{n} , \frac{i}{n} - s_i \right\}$. Extensive tables of significant values for D can be found in References [5], [9], and [12].

Finally, it is possible to combine the significance levels of the n-test and the k-s test, since under the null hypothesis, the significance levels attained by two independent statistics can be transformed into independent chi-squared statistics, combined, and then transformed back to an overall significance level. (See Appendix for details.)

Combining several truncated observation periods. If we have not one single observation period, but rather several, terminated by arbitrary events, can we still use the same test? Yes, if we simply put the different periods together end-to-end. It is as if we had a clock which was started from 0 when the first period began, and thereafter stopped whenever a period ended and restarted (without resetting) when the next period began. Under the null hypothesis of a Poisson process, the results expressed in terms of this clock time will look like a single realization. This can be seen most clearly from the second characterization of a Poisson process: as long as the clock is running, the probability of an event between (clock) time t and (clock) time t + h is the probability of an event in a certain small interval (of length h) which is part of one of the original periods, and thus is $\lambda h + o(h)$. (If a period ended between t and $t + h$, take a smaller h , so that the interval is entirely within a single period. This is valid, since the characterization is in terms of the limiting behavior as h approaches zero.)

Rearrangement of segments. Suppose one had 2k periods, each of length T , and all the odd-numbered periods were of one type while the even-numbered periods were of a different type. (For example, two modes of search could have been used alternately.) The null hypothesis is that the detection rate is the same for each of the two types of period. Now suppose one suspected that if the null hypothesis were not true, then one type of period would have a higher λ-value than the order, say $\lambda_1 > \lambda_2$. We wish to investigate the behavior of the test under this alternative. If the periods were put together as they occurred, then the graph of expected number of detections by time t would consist of straight line segments connected end-to-end, with slopes alternating between λ_1 and λ_2 , as illustrated in Figure 1. The K-S test (given the number of detections) compares the plot of actual number of detections by time t with a straight line whose slope is $n(2kT)^{-1}$, where n is the total number of detections. But n is likely to be close to its expected value,

$$\lambda_1(kT) + \lambda_2(kT) = \frac{\lambda_1 + \lambda_2}{2} 2kT ,$$

so that the slope of the line with which the sample plot is compared is likely to be close to

$$\frac{\lambda_1 + \lambda_2}{2} .$$

This line is also shown in Figure 1. It is obvious that the expected sample plot does not deviate very far from the reference line; thus it is likely that the actual sample plot will likewise not deviate very far. Now consider what happens if all the periods of type one are put together first, followed by the periods of type two. This situation is shown in Figure 2. The referance line is unchanged, of course, but the expected sample plot now deviates much farther. This procedure thus has greatly increased power against the suspected alternative and, of course, maintains the chosen size (because under the null hypothesis, the ordering of the periods is completely immaterial).

The situation just discussed is simple indeed, but the point is important. Because one can rearrange periods (or parts of periods) arbitrarily, without affecting the size of the test, one can choose a rearrangement to give good power against particular kinds of alternatives, not only in the simple case above, but in the other cases to follow. It is only necessary to make sure that the choice of breakpoints for rearranging does not depand directly on the actual location of events.

Multiple tests. It may happen that several different alternatives to the null hypothesis are under consideration. Nothing prevents the rearrangement and testing of the periods in several ways successively, one way for each alternative. The multiple-comparisons problem does come up, however, as discussed in Section 2. If the tests are not independent it is difficult to establish appropriate error rates. More will be said about this problem later.

4. Reduction of General Problem to the Poisson Case.

Re-statement of the problem. Several trials are conducted, the ith one lasting from time t_{1i} to time t_{2i} ; $t_{1i} \leq t_{2i} \leq t_{1\ ,\ i+1}$ for all i . For each trial, an entity known as a detection model is exercised to produce functions $\lambda_i(t)$; $\lambda_i(t)$ is defined for $t_{1i} < t \leq t_{2i}$. It is hypothesized that for any point t in the ith trial, the probability of a detection in a small time interval of length h

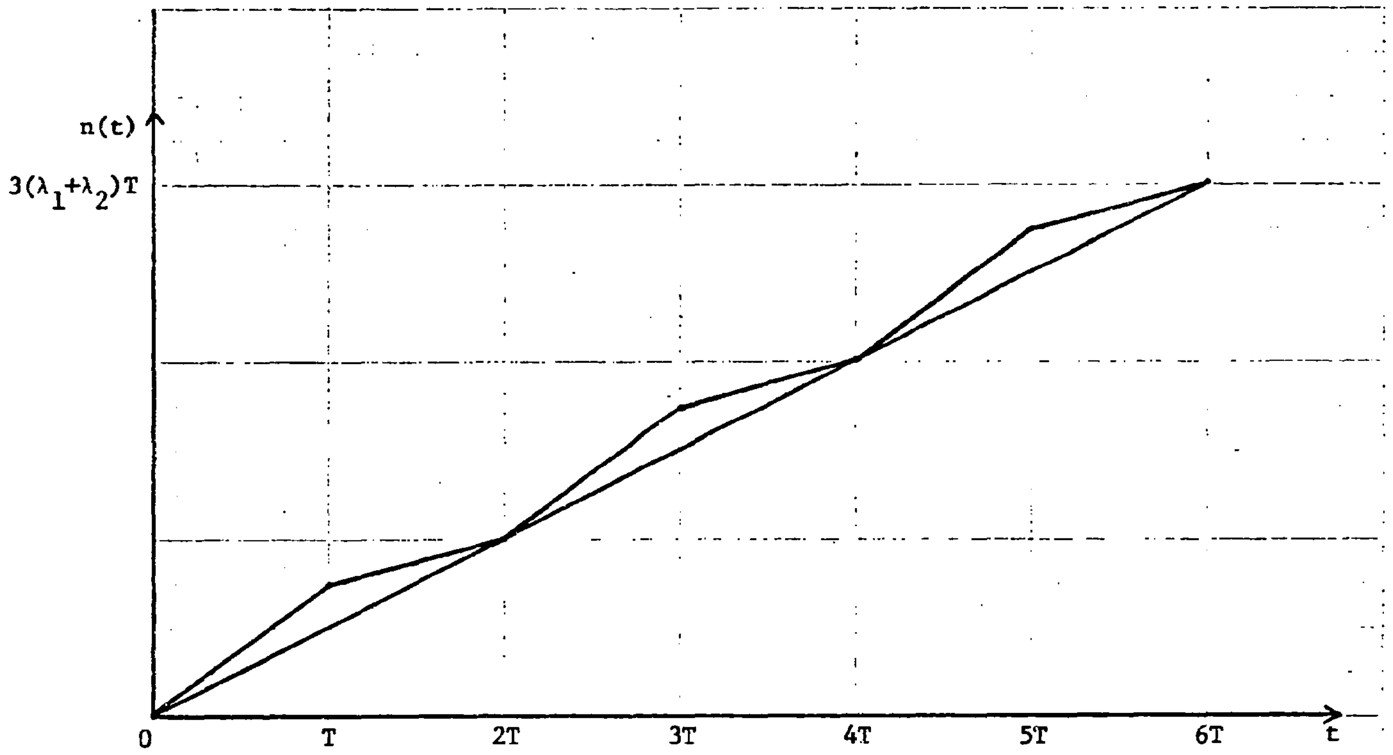


Figure 1. Expected number of detections by time t (top graph), and expected position of reference line (bottom graph). The line segments forming the top graph have slopes λ_1 and $\lambda_2 < \lambda_1$ alternately.

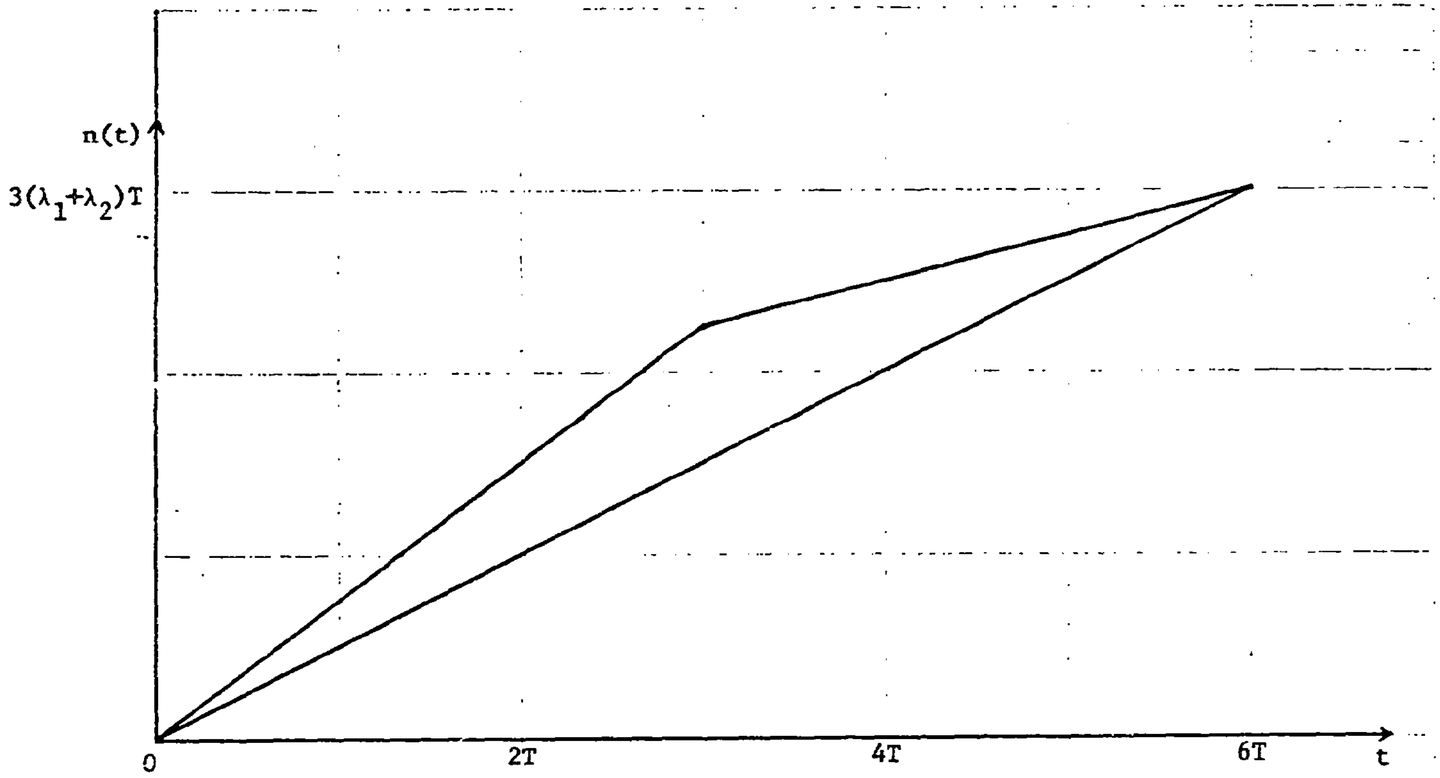


Figure 2. Expected number of detections (top graph), and expected position of reference line (bottom graph), with trials rearranged by type. The slope of the top graph is λ_1 for $t < 3T$, and λ_2 for $t > 3T$.

ending at t is $h\lambda_i(t) + o(h)$. A procedure is desired to test the hypothesis just given, such procedure to have a specified "size" α and a large "power" against certain reasonable alternatives to be specified below. Note that no restrictions are imposed upon the dependence of the $\lambda_i(t)$ on other variables, earlier times, or previous trials.

Implicit in the statement above is the condition that $\lambda_i(t)$ be finite for all realized t-values. This condition will generally be satisfied in practice. However, if it is not satisfied, the modifications necessary are simple. They have been relegated to the Appendix in order not to encumber the discussion at this point.

Reduction of the problem. To begin with, define a new function $\lambda(t)$ as follows: $\lambda(t) = 0$ for all t-values not in any of the trials; $\lambda(t) = \lambda_i(t)$ for all t-values within the ith trial. Note that $\lambda(t)$ is defined for all t, but is zero for all t greater than the time the last trial ended; it is also zero when no trial is in progress. Next, define another function $s(t)$ by the differential equation

$$\frac{ds}{dt} = \lambda(t)$$

and the boundary condition $s(t_{11}) = 0$. This new variable is essentially a new "time" variable: it can be thought of as the time read from a "clock" which begins operation any time before the first trial starts, and advances at a rate $\lambda(t)$. Another property of this clock is that at any time t, it registers the number of detections expected (under the given hypothesis) by that time. It does not advance between trials, nor after the end of the last trial. Being defined as the integral of a function which is defined (finite) and non-negative for all t, it is a continuous, non-decreasing function; and because $\lambda(t)$ is zero for all t beyond a certain value, $s(t)$ is a constant beyond that value.

Certain other variables germane to the problem are defined as functions of t, and will be denoted by $x_j(t)$; these include the input variables which are used by the detection model to produce $\lambda_i(t)$, and possibly other variables suspected of being important. Since we wish to continue the analysis of the problem in terms of s, we need to be able to express these variables as functions of s. Thus, we are led to consider an inverse function $t(s)$. Unfortunately, during the intervals when $\lambda(t)$ is zero, t varies but s does not. Thus, there is not a unique t value for certain s values, and one must be chosen. The obvious choice is the minimum value of t for each s, that is, where there are jumps in the graph of t vs. s, the value of t assumed at the jump is the lower value. This makes t a left-continuous function of s, so that the functions $x_j(t(s))$ are left-continuous functions of s, provided that the $x_j(t)$ are left-continuous. Now, it does not matter whether $x_j(t)$ is even defined at the instant a trial begins, because (with probability one) no detections will occur instantaneously; in fact, trials have been defined to exclude the starting time. However, it does matter very much that the functions be left-continuous at the end of a trial, to ensure that trials ending with a detection will have the proper values of the x_j associated with the time of detection. Thus, a definition which leads to left-continuity is a reasonable one to use; we henceforth take the $x_j(t)$ to be left-continuous.

Now, let us consider the problem in terms of s. Any given value of s corresponds to a value of t within (or terminating) some trial. To show that we have a Poisson process with detection rate unity, we will find it convenient to use the second characterization of the Poisson process in Section 3, and to consider short intervals running to the left rather than to the right (because of left-continuity). Thus, we ask whether the probability of a detection within a short interval $(s-h,s)$ is indeed $h + o(h)$.

The probability of interest is simply the probability of a detection within the original time interval $(t(s-h), t(s))$. Because $t(s)$ is left-continuous,

$$t(s-h) = t(s) - h\frac{dt}{ds} + o(h) = t(s) - [h/\lambda(t(s))] + o(h) =$$

$h + o(h)$ as required. We have prov.d that in terms of s, the cumulative number of detections is indeed a Poisson process with occurrence rate unity.*

The foregoing proof may seem too simple; one wonders about all the x-variables, for instance. It is important to remember that two distinct kinds of analysis are appropriate to a statistical testing procedure; the first, which might be called the null analysis, which assumes that the null hypothesis is true and then verifies that the test procedure does give a probability of rejection no greater than α; and the second, which examines what is likely to happen under various alternatives to the null hypothesis, and (hopefully) shows that the probability of rejection (the power) is high. What has been done above is entirely the first kind of analysis. Given the truth of the null hypothesis, we truly have a lot to work with. The detection model completely specifies the detection rate throughout each trial, taking account of all relevant variables, so that the cumulative number of detections is a stochastic process with independent increments and known incremental rate. This is an exceedingly simple situation, as statistical problems go, so it is not surprising that good tests are easy to come by once the problem is put into this framework. The next section considers the second kind of analysis, and opens up many more

*In case $\left.\frac{dt}{ds}\right|_{s-h} \to \infty$ as $h \to 0$, corresponding to $\lambda(t)\to 0$ as $t\to t(s)$, this argument does not work. However, it is easy to show (by integration of the elementary probability) that the number of detections in $(s-h, s)$ is a Poisson variable with mean h, from which it follows that the probability of one detection in $(s-h, s)$ is $h + o(h)$, and the probability of more than one detection is $o(h)$.

questions; it is concerned, one might say, with the art part of the testing problem. This section concludes by pointing out that one can break up and rearrange segments of the plot of number of detections against s at will, as discussed in Section 3, and still apply the standard tests for total number of detections and for distribution ot detection s-values; one can consider all the x-variables in doing this rearrangement; and one can do it several different ways, with appropriate consideration of the error rates. It must be said, however, that one cannot be completely arbitrary in rearranging things. For example, suppose one chose to take a very small interval around each discontinuity of one of the x-variables, say range, and put all these intervals first; then append all the remainder. If the interval length is chosen small enough, all the chosen intervals will amount to (say) one percent of the total s-length; however, since any detection terminates a trial and the next one will (almost surely) start with a different range value, all the detections will be found in that one percent! Thus, the test will reject even a correct model with certainty if carried out this way. On the other hand, if the whole s-length is broken into small segments, and these arranged in order of range, the test is valid (see the Appendix for a proof). The criterion that must be satisfied is that under the null hypothesis, knowledge of the basis of rearrangement does not affect the expected pattern of detections in s . This is obviously not satisfied when the discontinuities are gathered first, and, intuitively, is satisfied when discontinuities are ignored, since all the relevancy of all the variables is assumed to be correctly handled by the model.

5. Practical application of methods outlined above

Choice of tests. The objective of this whole procedure is to test the adequacy of a detection model which purports to correctly calculate the (instantaneous) detection rate in terms of certain input variables. It has been shown in previous sections that the s-range can be split up into segments and rearranged almost at will, with the implication that different arrangements will be good for different things. This section will consider rationales for different possible arrangements, practical questions related to carrying out the tests, and statistical considerations involved in multiple tests.

Even with no reordering, the test has some power against any alternative to the null hypothesis. However, as illustrated in Section 3 (see Figures 1 & 2), reordering can greatly increase the power. The problem is to choose from the many possible orderings a reasonably small number which will (together) have high power against a reasonable range of alternatives.

Let the set of (possibly) relevant variables be denoted by $\{x_j(s)\}$. These will include the variables which serve as input to the model, and quite possibly others (e.g., operator identification, time of day, -r even the number of trials since the last successful one). If the model is not correct, then its deficiencies should show up as inadequate or incorrect treatment of one or more of the $x_j(s)$. Thus, it makes sense to pick out the most likely candidates among the x_j , and rearrange the data in order of one of these variables at a time; then look at a plot of number of detections versus (rearranged) exposure s , and use the Kolomogorov-Smirnov test to see if the detections are appropriately distributed. For instance, if it is suspected that the model's treatment of target speed is inadequate, arrange the s-values in order of increasing target speed, and then plot the cumulative number of detections; it should look like a sample uniformly distributed over the s-range. The rationale for plotting in increasing values of the suspect variable is that the segments having values of that variable for which the model overestimates the detection rate are put together, as are the segments for which the model underestimates the detection rate, thus tending to maximize the K-S statistic and the power of the test. If it should happen that our suspicions are sufficiently definite, we might do better yet. For example, if it were felt that the model underestimated both for small values and for large values of the variable, but overestimated for intermediate values, there is no reason why we could not put the large and small values first and the intermediate values last. For another example, it may be that two variables interact in such a way that their quotient, or sum, or some other function of the two, determines whether the model overestimates or underestimates. If one suspects a certain kind of interaction, then one can rearrange accordingly, taking account of both variables together.

And again, suppose that one suspects a variable which depends not only on the current time but also on the past -- as for instance the length of time that a signal has been above a certain strength, or the integral of a variable over some interval ending at the present. All that is required is that that variable be determined for each time segment, before the segments are broken up and rearranged (since it cannot be determined once the ordering is changed). It is then available to use in exactly the same fashion as any other variable. Finally, one can do a one-sided K-S test instead of a two-sided test in some circumstances, thus increasing the power. This can be done when it is known whether the sample curve will tend to be above the reference line or below it -- i.e., when we know not only which segments are likely to deviate in the same direction, but also which direction it is. The test statistic is either

$$\max \{s_i - (\frac{i-1}{n})\} \text{ or } \max \{\frac{i}{n} - s_i\} \text{ ;}$$

it is referred to the same table to test significance, but at double the desired error rate -- i.e., a 5% test is performed using the (two-sided) 10% value.

As a practical matter, one could select (say) the five most suspect variables, and do a plot for each. As long as one does not plot too many variables, and the variables are unrelated, one might reasonably assume that the resulting tests are independent. Thus, to achieve an overall 5% error rate (size), each of the five tests would have to be done at the 1% level: $(1 - .05) \cong (1 - .01)^5$. Note that the case of both extremes alike and the middle different, mentioned above, would show in this procedure as a sample plot which has an S-shape relative to the reference line.

Performing the test. Consider, now, the realistic situation wherein trials have been run, and the data are stored in digitized form on magnetic tape, ready for the computer. What should we do?

Notice that the recording in digital form at times t_1 , t_2, ..., implies a segmentation of the trials into time segments Δt_1, Δt_2, ..., not necessarily equal in length. (Of course, the s-length corresponding to a segment is not usually equal to the t-length, anyway.) But then, we do not need to have segments of equal s-length; we only need to have them short, and to know the s-length. Remembering the definition of s , we have (to within the approximation implied by the digitization)

$$\Delta s_i = \lambda(t_i)\, \Delta t_i \text{ , where } t_i$$

is the time that goes with the segment Δt_i . Also, we assign a value s_i , calculated as $s_i = \Delta s_1 = \Delta s_2 + \ldots + \Delta s_i$.

At this point, we have the data in the form of records, one for each segment, each consisting of values of s , Δs , all the x_j , and an indication of whether a detection occurred (presumably at the end of the segment involved). It is now straightforward to plot number of detections versus s , or to rearrange the segments according to the values of some of the x_j and then plot number of detections (not against s , which no longer will be an increasing function of position in the list, but against the sum of the Δs values as we proceed through the list). Two tests are done: first, comparing the total number of detections with its null hypothesis distribution (a Poisson distribution with mean value equal to the sum of all the Δs values), and second, calculating a Kolomogorov-Smirnov statistic to test whether the detections look like a sample from a uniform distribution across the (rearranged) s-values.

The Poisson test is trivial: sum all of the Δs values to get the mean value, m ; pick an error rate, α_1 ; then go to the tables of the Poisson distribution for the calculated value of m and see whether the actual number of detections is within the appropriate (two-tailed) limits. A Poisson table can be found in Reference [9]; more extensive tables can be derived from Chi-squared tables or tables of the incomplete Gamma function, since if X is Poisson with mean m , $P(X \leq k) = P(T_{k+1} > m)$ where T_{k+1} is the sum of $k+1$ exponential variables, each with mean 1 ; T_{k+1} has the Gamma distribution with parameter (mean value) $k+1$, so that $2T_{k+1}$ has a Chi-squared distribution with parameter $2k+2$. Chi-squared tables are in almost any statistics text; the incomplete Gamma function is tabled in References [1] and [10].

The K-S test is almost as simple: plot number of detections against the running sum of Δs values; connect the end point of the plot to the origin by a straight line; find and measure the maximum (absolute) deviation between these two, and divide by the number of detections to obtain the fractional deviation; choose an α-value, α_2 , and compare the fractional deviation with the tabled values to check significance. A look at the example will make clear the procedures, including the modifications for a one-sided test. Tables can be found in, e.g., References [5], [9] and [12].

Computer efficiency

The techniques just given are conceptually simple. However, there is a simple pitfall which may make them surprisingly costly to implement, unless the implementation is properly done. The problem is the number of data records that exist, and the sorting operation necessary to rearrange them in order of some particular x-variable. If each segment represents (say) five seconds, and twenty trials last an average of (say) two hours, there are 28,800 data records stored. Sorting these records is time-consuming. Perhaps a few tests can be done with no problem, but, if the techniques are to be used as diagnostics, necessitating many different plots, then more efficient techniques would be very useful. One approach will now be given.

Suppose each record involving a detection were duplicated into a separate list. (There will be rather few of these, no more than the number of trials, of course.) Let n be the number of detections. Now, if it is desired to plot against (say) x_2, one need only order the x_2-values in this list and then make one pass through the entire set of data, adding each Δs value into the proper one of $n+1$ registers, the first for those records having an x_2-value less than the smallest of the n ordered values, the second for values between the smallest and the next-smallest, etc. Once this operation is performed, one has all the information necessary to plot the appropriate sample curve. In fact, one could just as well let the computer find the maximum deviation, since it will occur at one of the n steps in the plot. Denote the totals in the registers by S_0, S_1, ..., S_n, and the total of these by S. The ordinates of the reference line at the appropriate values are, of course, S_0/S; $(S_1+S_0)/S$; ... ; $(S_{n-1}+\dots+S_0)/S$. The corresponding ordinates of the sample plot are, of course, $1/n$, $2/n$, ..., $(n-1)/n, 1$. The maximum absolute difference is, thus,

$$\max_{1\le i\le n}\left\{\frac{S_0+\dots S_{i-1}}{S}-\frac{i-1}{n}\ ,\ \frac{i}{n}-\frac{S_0+\dots+S_{i-1}}{S}\right\},$$

which can easily be calculated by the computer. For one-sided tests, the statistic to be calculated is

$$\max_{1\le i\le n}\left\{\frac{i}{n}-\frac{S_{i-1}+\dots+S_0}{S}\right\}$$

if the sample plot is not expected to be below the reference line, and

$$\max_{1\le i\le n}\left\{\frac{S_{i-1}+\ldots+S_0}{S}-\frac{i-1}{n}\right\}$$

if the sample plot is not expected to be above the reference line.

6. Diagnostics, model-building, and other extensions

This section presents a short discussion of topics beyond the scope of this report: the possibility of modifying the techniques presented in this report so that they will serve a diagnostic function, identifying weaknesses in the model more specifically; using a regression approach with these techniques to suggest a better model; and extension to cases where the "time" variable is essentially discrete in nature. Anything written here is, of course, preliminary, and therefore subject to modification upon further investigation.

Refer to Figure 17, or even better, to Figure 8, in the next section. If the model were correct, the step function would closely follow the diagonal. Since this is not the case, it is fairly clear that the model's treatment of range leaves something to be desired, unless there is another variable which just happens to be correlated with range in this set of trials. Of course, the structure of the trials is very simple in this case, so not too much can be expected in the line of involved trouble-shooting. Where there are more variables, and they occur in various combinations, there will be more opportunity (and more need) for careful scrutiny to determine which variable needs correction. The idea here is not to insist on formal statistical tests with strict significance levels, but rather to try different things, in order to get ideas about the model which will lead to further investigations -- either by more full-scale trials, or by theoretical means, or by specific tests aimed at verifying assumptions or small parts of the model. For example, it might be that there is a sub-model for cloud diffraction of radar waves which is one of the foundations of a model for signal strength. If it turned out that the range variable seemed suspect, but mainly when cloud cover was unusual, one might be led to look carefully at that particular sub-model with an eye to specifying experiments which would verify or modify it as needed. Note that this would require more than simply looking at plots: it would also require comparing those areas on the plots where agreement was not good, in terms of other variables not directly involved in the plot under consideration. Multi-dimensional plots might be better, but unfortunately they become hard to interpret in more than two dimensions, and also the amount of data might not be sufficient to enable visual interpretation. This leads naturally into the second subject of this section, namely regression techniques (and related techniques, such as correlation).

The interpretation of masses of data is sometimes facilitated by statistical techniques such as regression or correlation. One could think of the experimental results as a very large number of trials, one for each little time segment, most of which resulted in no detection. One wishes to analyze the occurrence of detections, in terms of the available explanatory variables. This amounts to fitting a "response surface", where the dependent variable is 1 or 0 (detection or nondetection), and the independent variables are the explanatory variables just mentioned. The fitted surface will be a representation of the instantaneous probability of detection (the detection rate). With this technique, it may be possible to get around the problem of spurious significance caused by correlation between different explanatory variables, since one can look at the effect of one variable with another "held fixed" (in a statistical sense). The problem is not as simple as classical multiple regression, however, because the dependent variable takes on only two values, 0 and 1 , and most of the time it takes on the value 0 . It seems likely that a technique of smoothing (of the response variable) is required before any fitting is attempted. There are several possibilities for such treatment, but none have been seriously investigated by this writer.

Finally, a few words are in order about the situation wherein the time variable is essentially discrete. For example, suppose detailed records are not kept during a trial, so that one knows only the overall probability of detection for a long segment of time, as well as whether detection actually occurred. Of course, one cannot break up the segments to rearrange by values of a given variable, since one does not know the values of that variable within the segment. However, one may have the values of some meaningful variables, so that it may still make sense to ask about rearrangements of trials. Can it be done? Essentially, the answer is yes. We have many binomial trials, with differing probabilities of success. This does not fit the Kolmogorov-Smirnov framework directly, but it may be forced into it, possibly in more than one way. One possibility is to manufacture a Poisson process out of it, thus arriving back at the situation already analyzed. Another may be through a modification of K-S, since for an arbitrary ordering of trials, a plot of number of detections by trial i against the sum of the probabilities of detection up through trial i will tend (under the null hypothesis) to follow the diagonal, just as if the probabilities were identical for each trial. (The K-S statistic is applicable to the limiting case wherein the probabilities for each trial go uniformly to zero while the number of trials goes to infinity; our case could be treated by K-S if we could find an error bound to allow for the discontinuity.) At the moment, the former method looks preferable. It seems likely that one could use that approach, with only minor losses due to the artificial imposition of the Poisson process.

7. An example

Suppose one sets out to test a detection model which involves just two variables: range and power. Suppose the hypothesized model is the following: $\lambda(t) = kp\, r^{-2/3}$, where k is a constant, r is the range and p is the power. A certain number of trials are scheduled, for each of which the target starts at range 60 km and moves directly over the hunter, at a constant speed of 2km/hr; the trial is terminated at detection, or at time 40, whichever happens first. Two values of power are used alternately, so that odd-numbered trials use the value p_1, even-numbered trials use the value p_2. Thus we can write $\lambda(t) = kp_i\, |60 - 2t|^{-2/3}$, where $i = 1$ or 2 according as the trial number is odd or even. Suppose finally that the constant k is hypothesized to have the value 0.6, and $p_1 = .03$, $p_2 = .06$. Now it can be calculated that the probability of a detection somewhere within a trial is approximately 1/2, so that we expect to have detection in about half of the trials.

Rather than go out and build a physical system to satisfy the conditions just specified, we choose to simulate the system. But to do so, we need to know the true situation - i.e., the "true model" - so that we can determine when detections actually happen. Let us use for a true model the function

$$\lambda(t) = p_i\, |60 - 2t|^{-1/3} ,$$

i.e., k is set to unity, the exponent is changed, but p_i is left alone. We will use this model to find when (if at all) detection takes place for each trial; then we will analyze the trials as if we did not know the true model, to see how well the test procedure works.

First we need to calculate times of detection. Let $P(t) = \Pr(\text{no detection by time } t)$. Then the function $P(t)$ satisfies the equation

$$P(t+h) = P(t) \cdot (1-h\lambda(t))+o(h) , \text{ for small } h .$$

Easy manipulation gives

$$\frac{P(t+h) - P(t)}{P(t)} = h\lambda(t) + o(h) ,$$

or

$$\frac{d}{dt} \log P(t) = \lambda(t) ,$$

or finally

$$P(t) = \exp - \int_0^t \lambda(u)du ,$$

where the (multiplicative) constant of integration is set equal to unity because $P(0) = 1$. Now define a function $G(t) = 1 - P(t)$ for t less than t_{max} , and $G(t) = 1$ for t greater than or equal to t_{max}. This $G(t)$ is the cumulative distribution function (cdf) of the termination time of the trial. Since any proper cdf is itself a random variable having a uniform distribution on the unit interval, one obtains random values of t by simply drawing random numbers, equating them to $P(t)$, and solving the resulting equation for t .

In summary, then, one finds the proper value of t for the ith trial by equating a random number u_i to $P(t_i)$ and solving for t_i ; if t_i is less than t_{max} , it is the time of detection, and otherwise, no detection occurs on that trial.

Of course, the $\lambda(t)$ used in this calculation is the true $\lambda(t) = p_i \left|60-2t\right|^{-1/3}$: Integration of this formula gives the following result: t_i is the solution of

$$u_i = \exp - \frac{3\ p_i}{2} \left[60^{2/3} - (60 - 2t)^{2/3}\right] \quad \text{for} \quad t \leq 30 ;$$

$$u_i = \exp - \frac{3\ p_i}{2} \left[60^{2/3} + (2t - 60)^{2/3}\right] \quad \text{for} \quad t > 30 .$$

Only one of these two equations will have a solution for any u_i . Of course, as mentioned before, t_i is set equal to t_{max} if it turns out to be greater than t_{max} .

Figure 3 is a graph of both true and hypothesized $\lambda(t)$, for t-values between 0 and 40 . Figure 4 is a graph of the cumulative probability of detection, i.e., of Pr(detection by time t), for the same range of t , and for the low value of p_i ; Figure 5 is the corresponding graph for high p_i . It can be seen that the average probability of detection is close to 1/2 , for both the true and the hypothesized models.

Several sets of random numbers (RN's) were chosen to test the techniques presented above. For each set, three figures are presented: The first gives a plot of cumulative fraction of detections against expected fraction of detections, for trials in natural order (as drawn); the second is a similar plot, but with all the low-p_i trials first, followed by the high-p_i trials; and the third gives the plot in range order as described earlier, that is, plotting actual number of detections obtained at ranges less than a given value against the expected number of detections for the same interval of range values. (For the illustrative purposes of these examples, a special-purpose program was written which takes advantage of the fact that only two different kinds of trials were considered, with the λ-values for one being exactly twice the values for the other. It is simply a messy bit of bookkeeping, so will not be described here.) The first set consisted of 100 RN's, and led to the results in Figures 6 through 8. Because of the relatively large number of trials, these curves are rather smooth, and give a good representation of the "expected" behavior of the method. Next, a set of 30 RN's was chosen. Unfortunately, this set happened to be rather nonrepresentative, in that the odd-numbered RN's tended to be larger than the even-numbered ones. (The odd-numbered ones had a mean that was 2 sigmas above expectation, while the even-numbered ones had a mean 1 sigma below expectation.) This led to a wiggly behavior for the first plot (Figure 9), and a spurious but definite effect for the second (Figure 10); the effect was considerably more significant on the third plot, Figure 11. (Remember that the model utilizes the p_i correctly, up to a constant factor, so that any effect that shows up on the second plot is either completely spurious or is caused by a sort of correlation between the p_i and the range variable.) Accordingly, the same set was used again, but with the odd- and even-numbered subsets interchanged. This led to Figures 12 to 14, wherein the same sort of behavior is evident on the first, a reverse behavior on the second, and less significance on the third. Here the chance inadequacies of the set of RN's tended to counteract the real effect of the range variable. Then a fourth run was made, using another set of 30 RN's for which there was no anqmalous behavior; these results are presented in Figures 15 to 17, and show the kind of behavior to be expected in general.

On each plot, the value of the Kolmogorov-Smirnov statistic indicating departure from uniformity, and its significance level, has been printed. It is clear that the rearrangement of trials has greatly increased the significance level of the test.

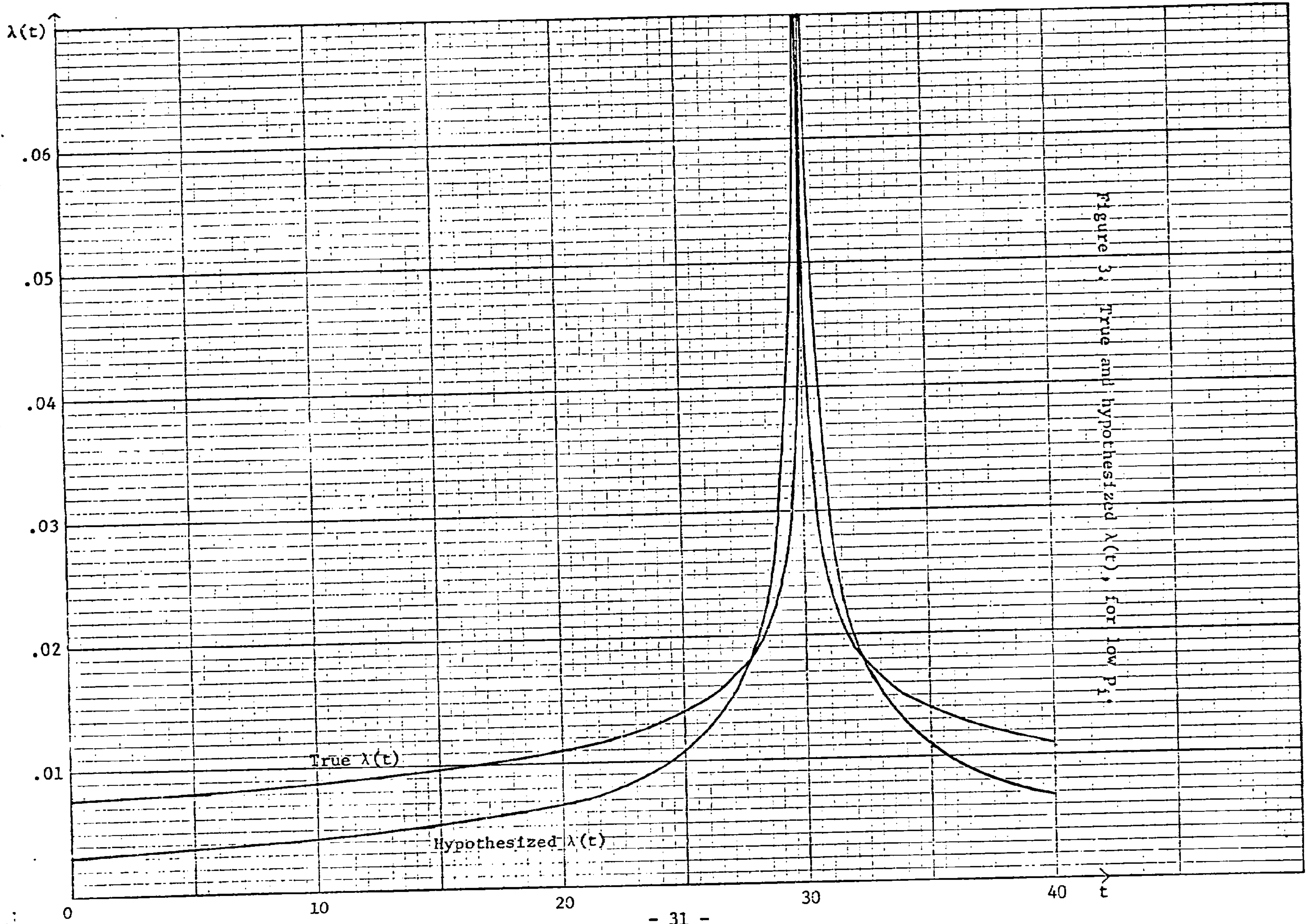
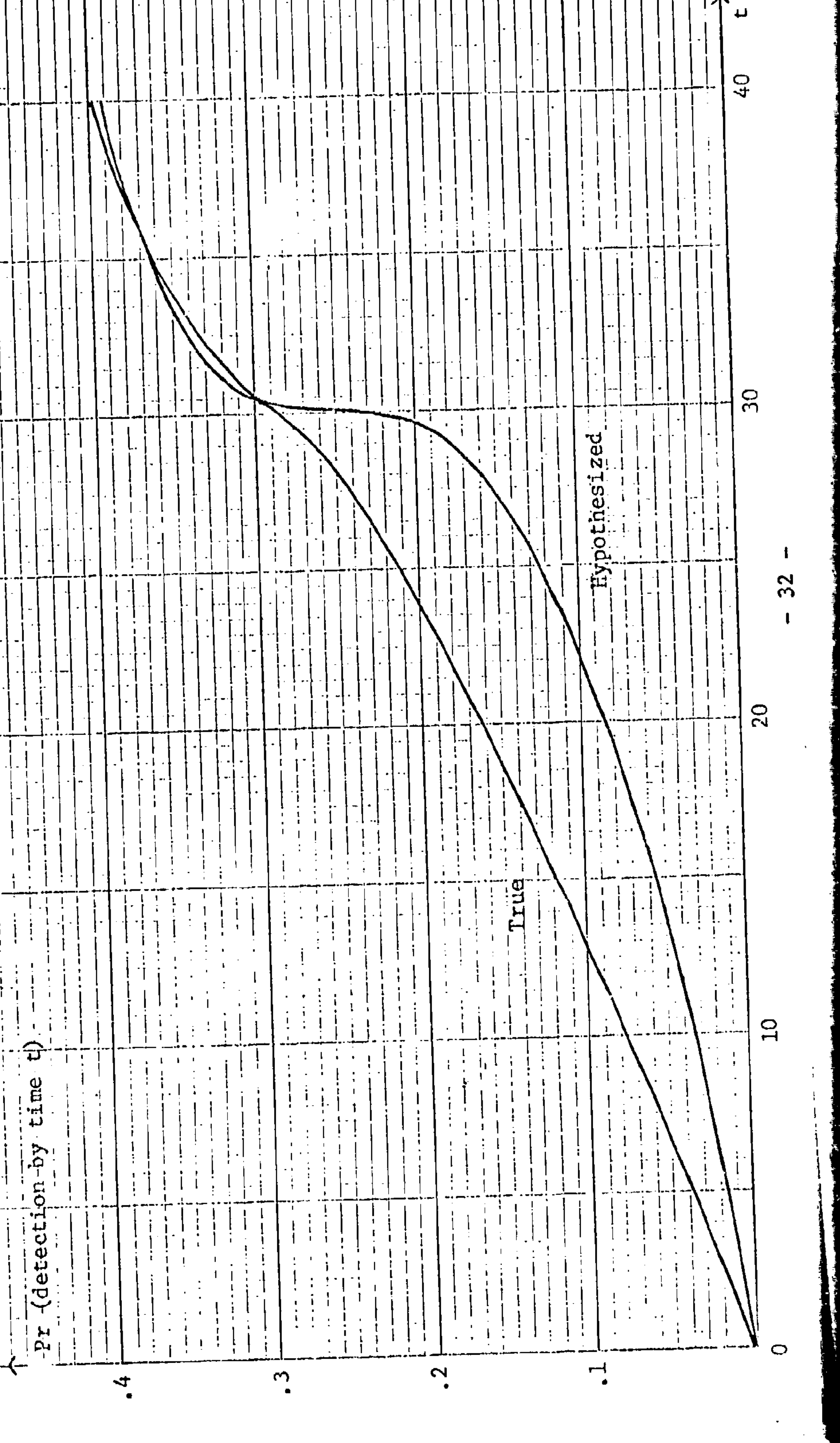


Figure 3. True and hypothesized λ(t), for low p_i.

Figure 4. Cumulative probability of detection by time t, low p_i.



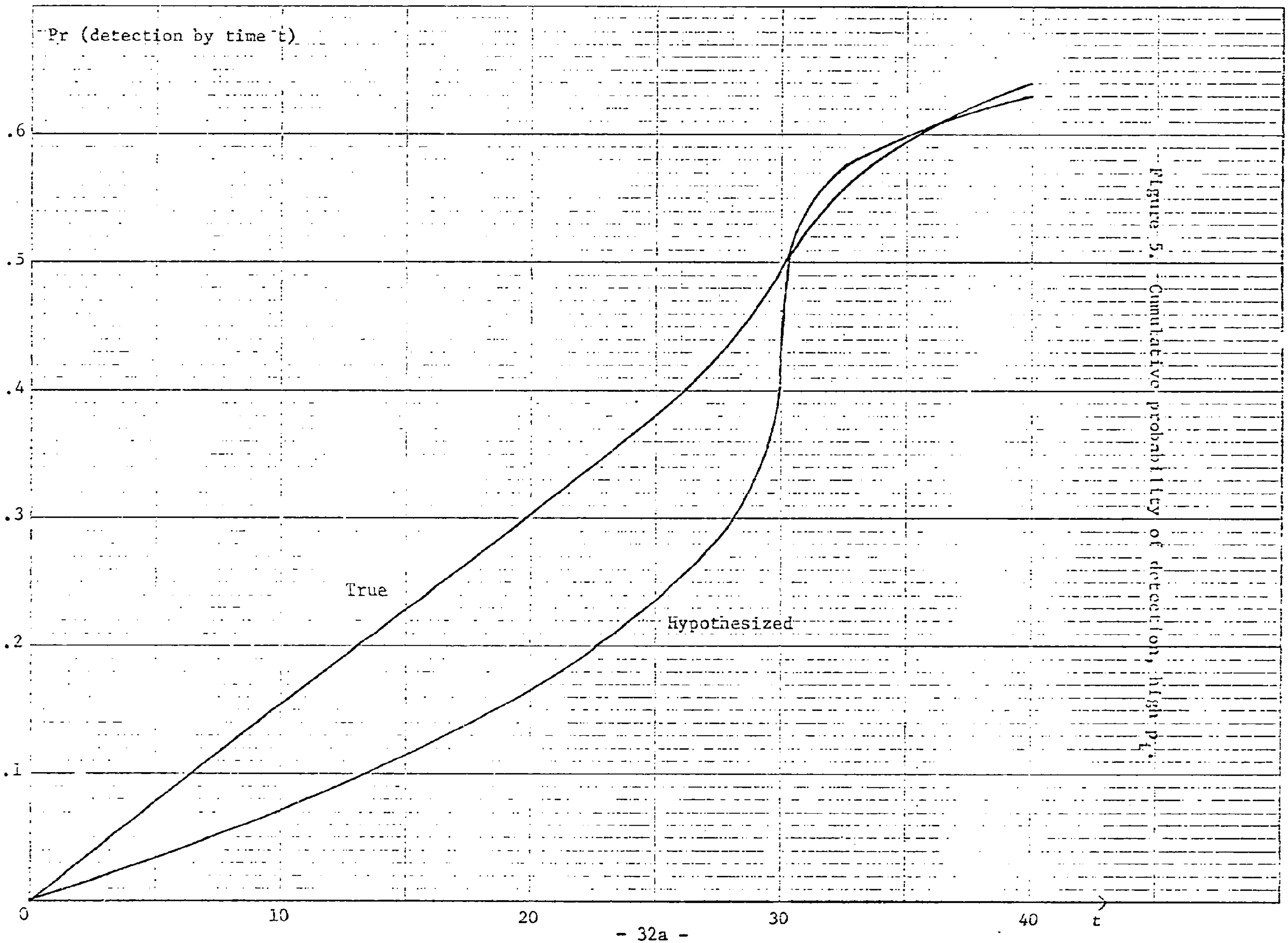


Figure 5. Cumulative probability of detection, high p_i.

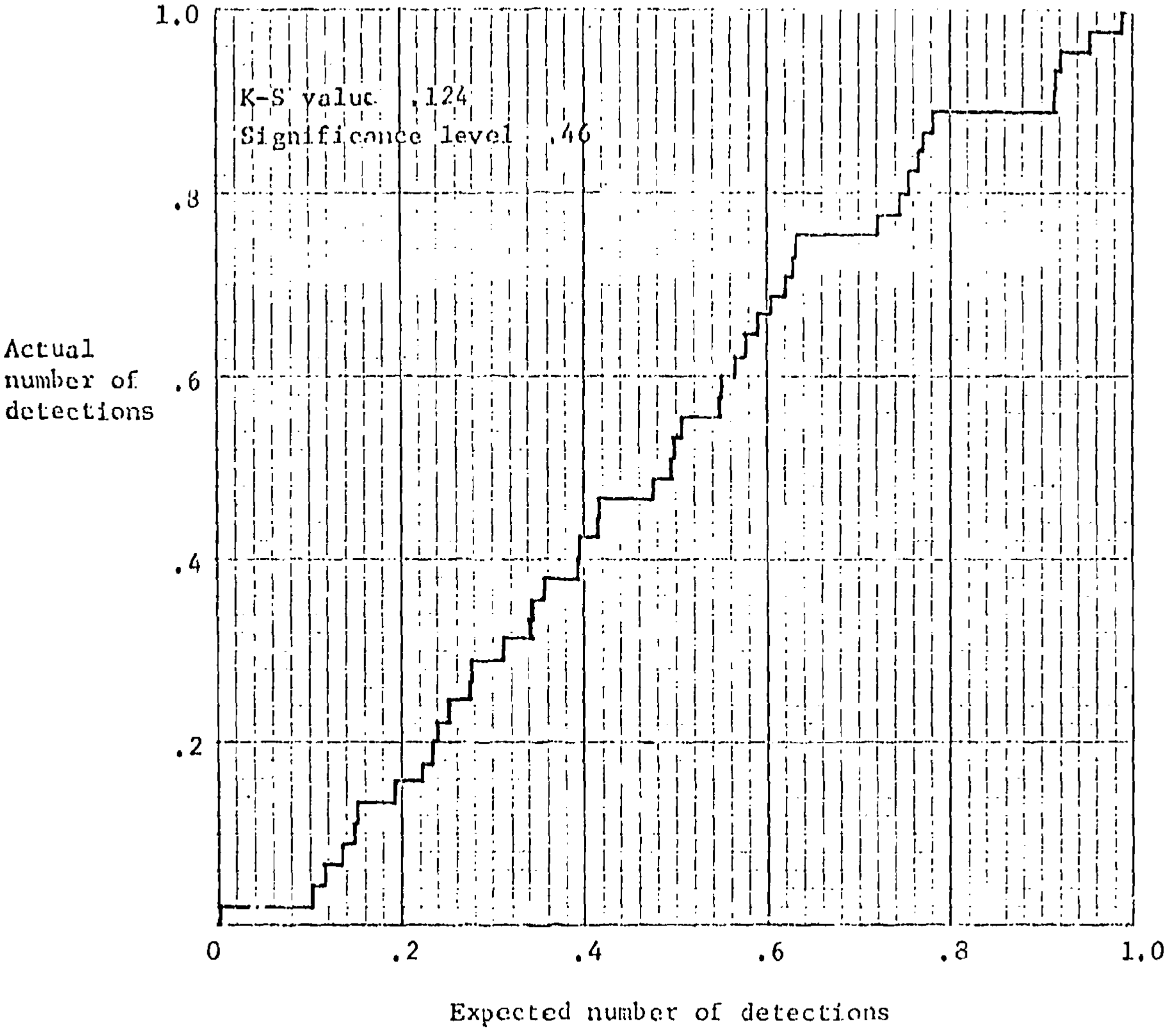


Figure 6. Plot of actual vs. expected number of detections, 100 trials, natural order.

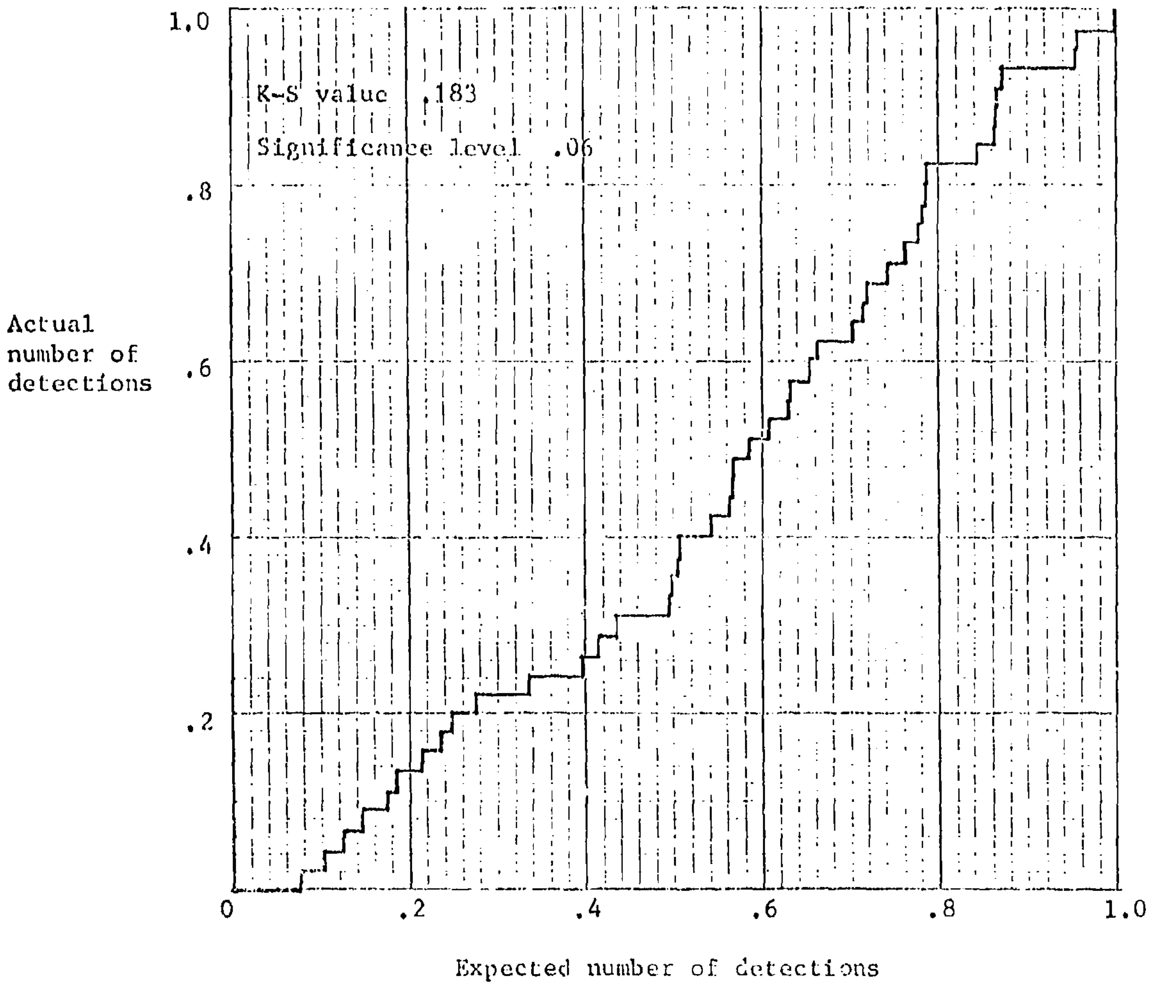


Figure 7. Plot of actual vs. expected number of detections, 100 trials, p_i order.

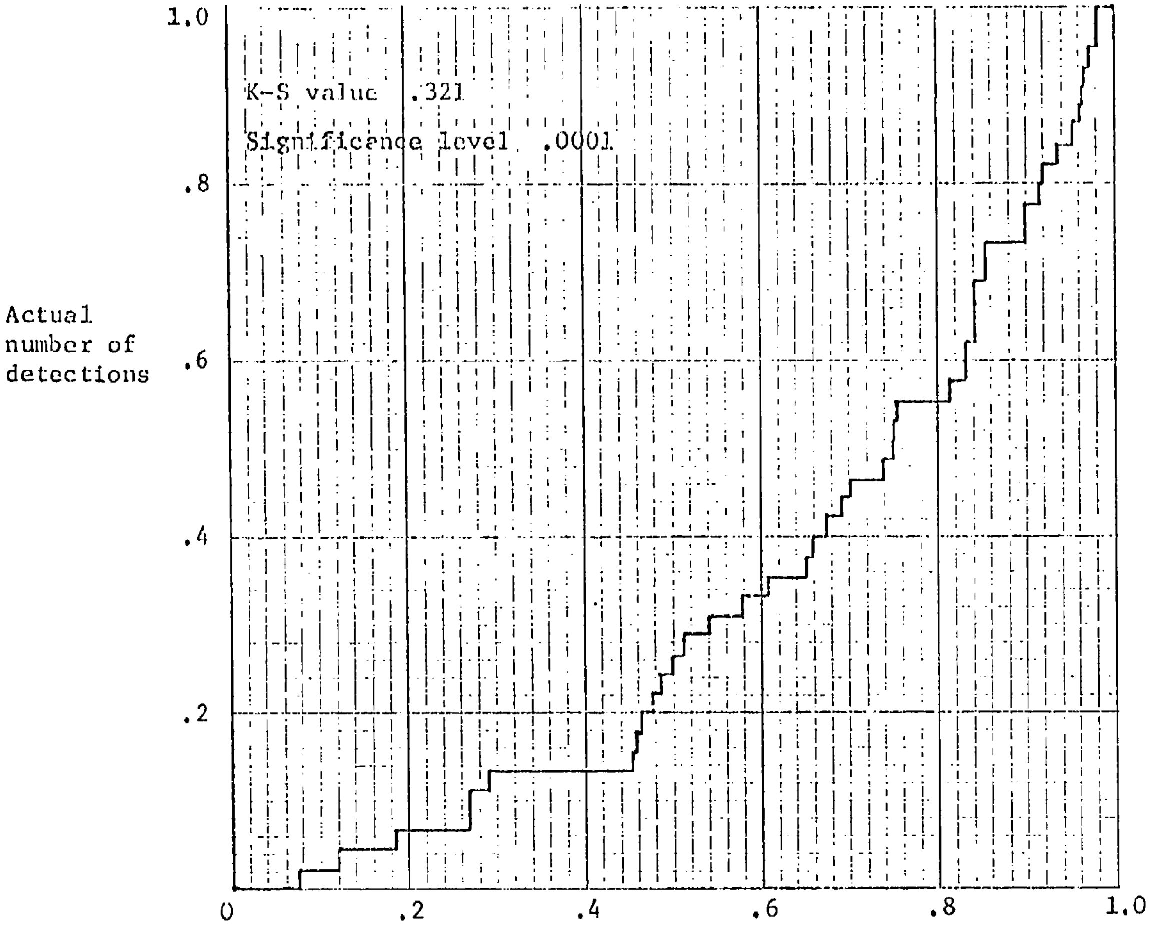


Figure 8. Plot of actual vs. expected number of detections, 100 trials, range order.

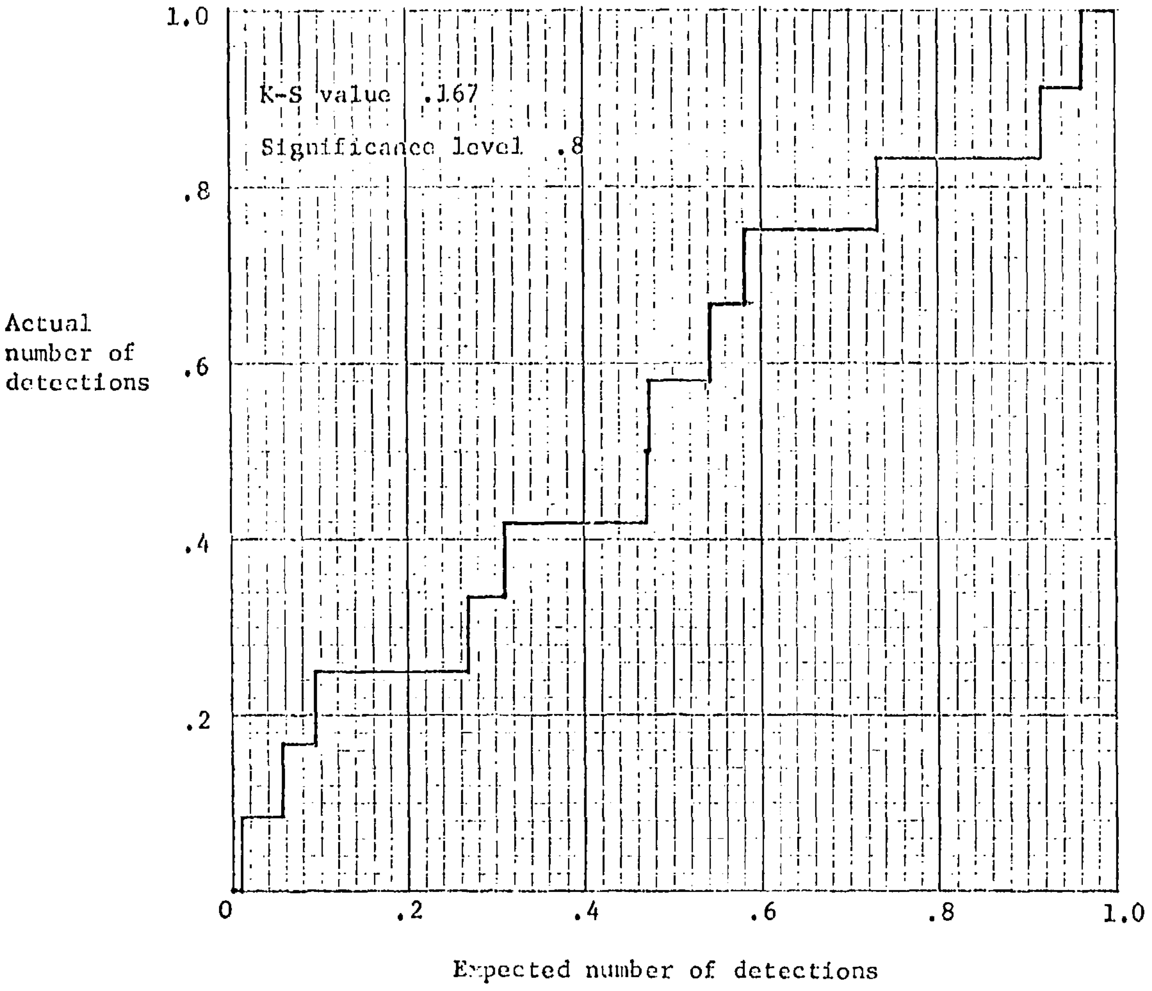


Figure 9. Plot of actual vs. expected number of detections, 30 trials, set A, natural order.

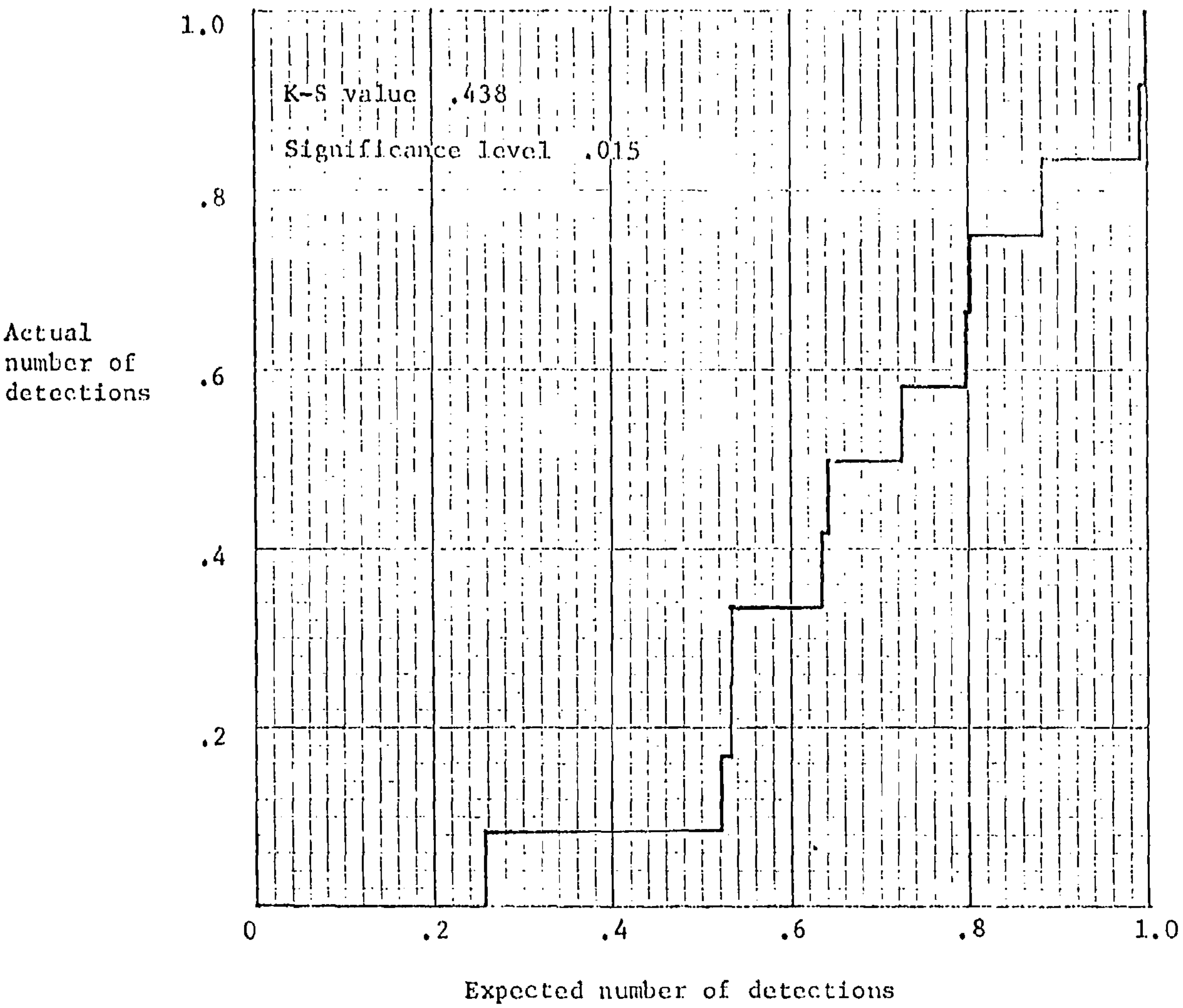


Figure 10. Plot of actual vs. expected number of detections, 30 trials, set A, p_i order.

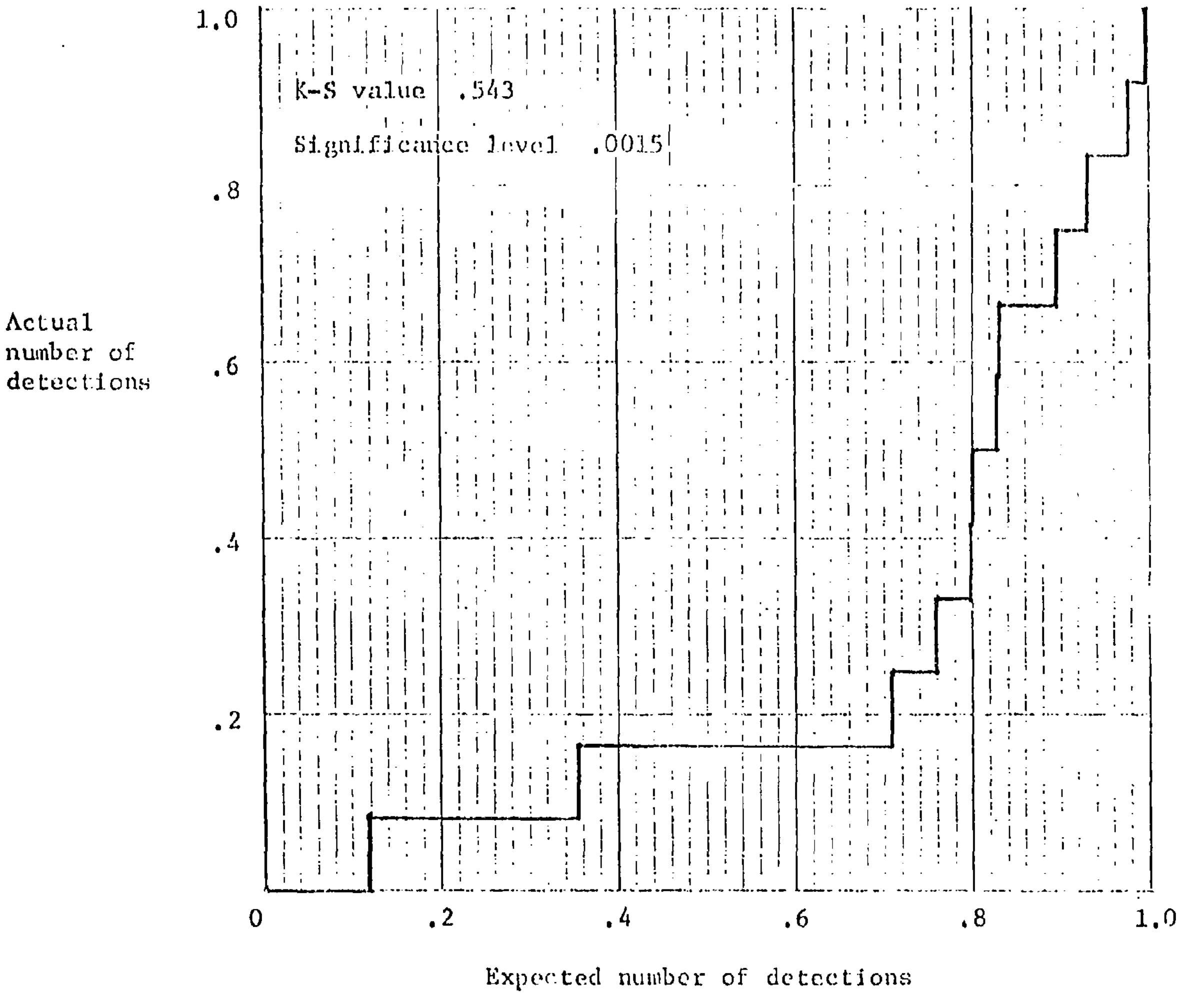


Figure 11. Plot of actual vs. expected number of detections, 30 trials, set A, range order.

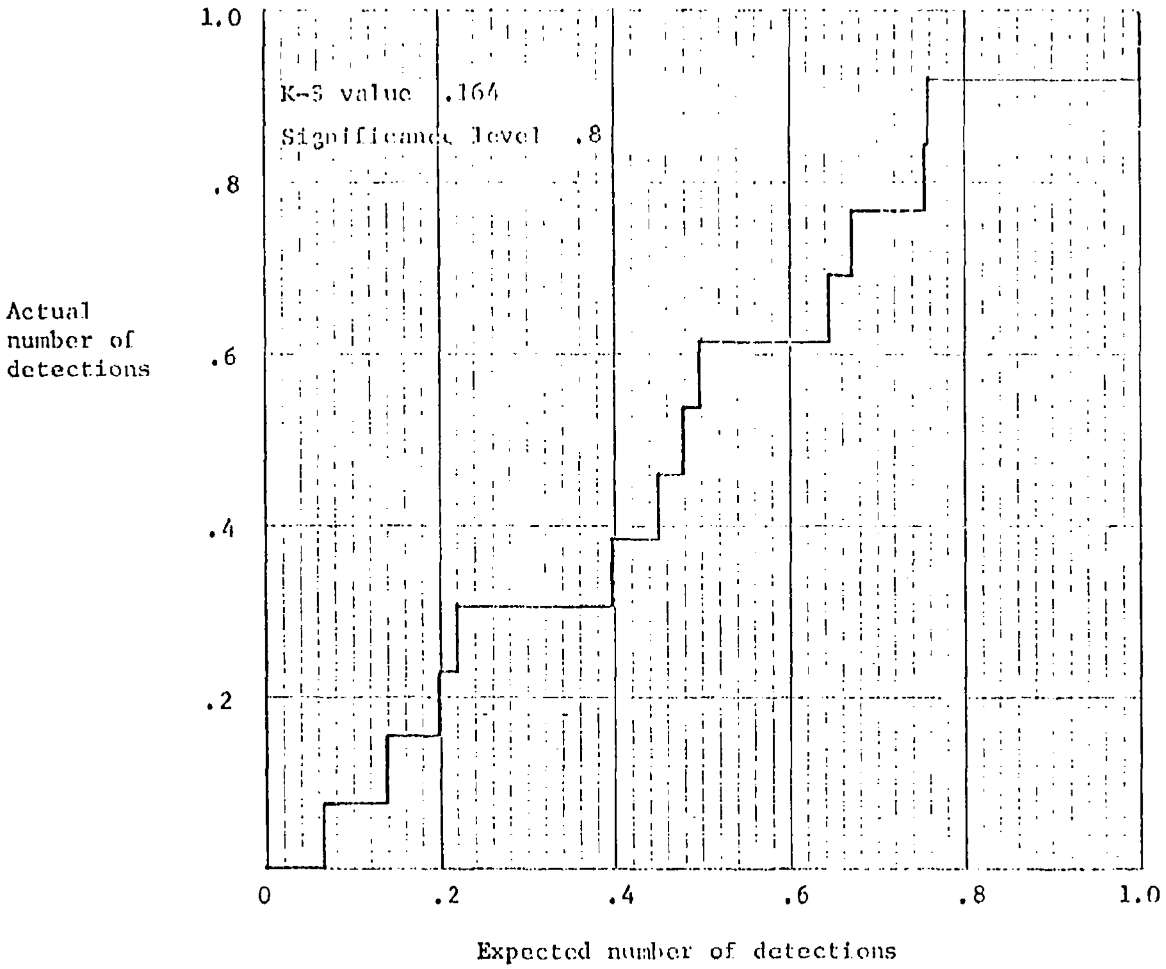


Figure 12. Plot of actual vs. expected number of detections, 30 trials, set A (reversed), natural order.

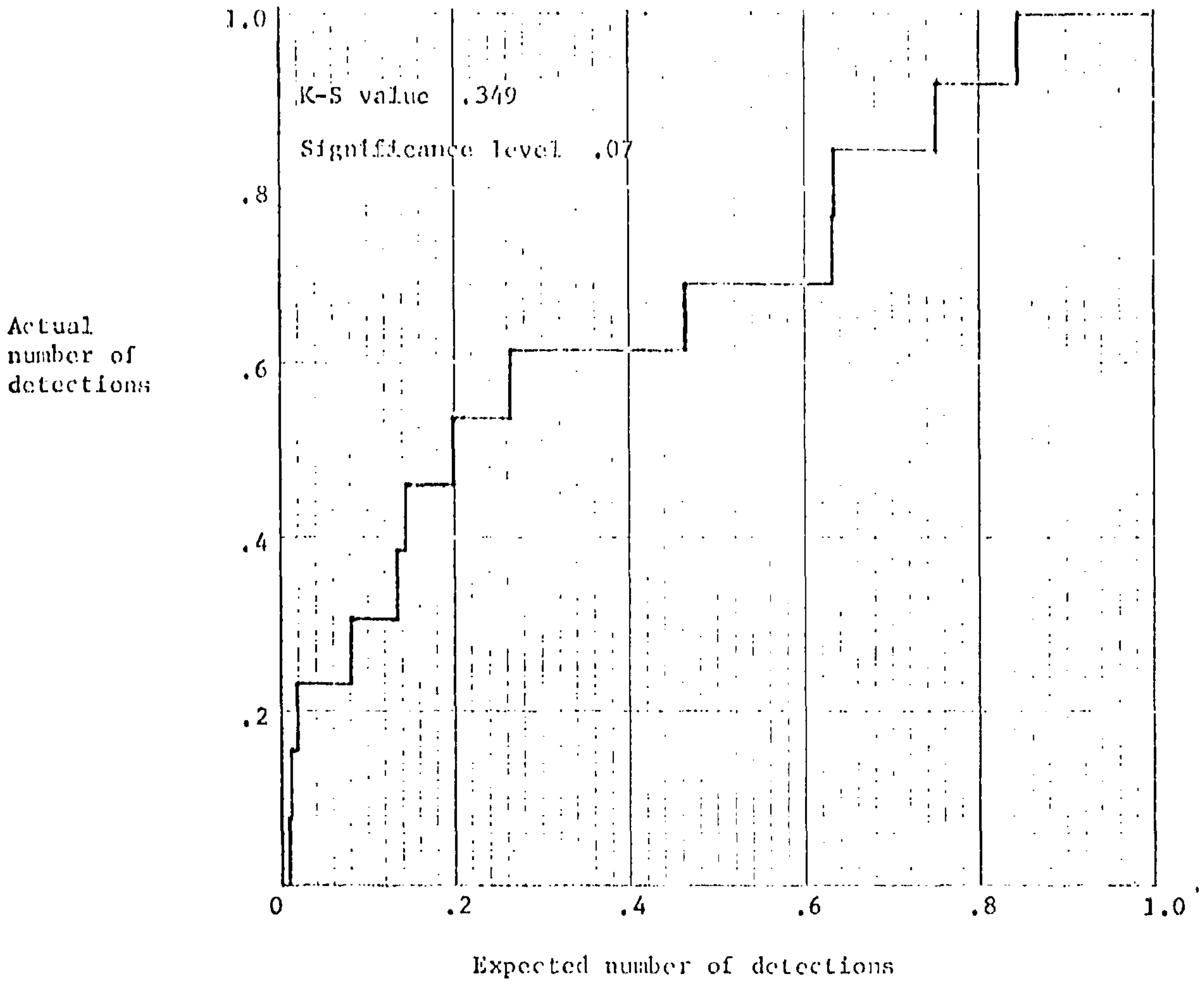


Figure 13. Plot of actual vs. expected number of detections, 30 trials, set A (reversed), p_i order.

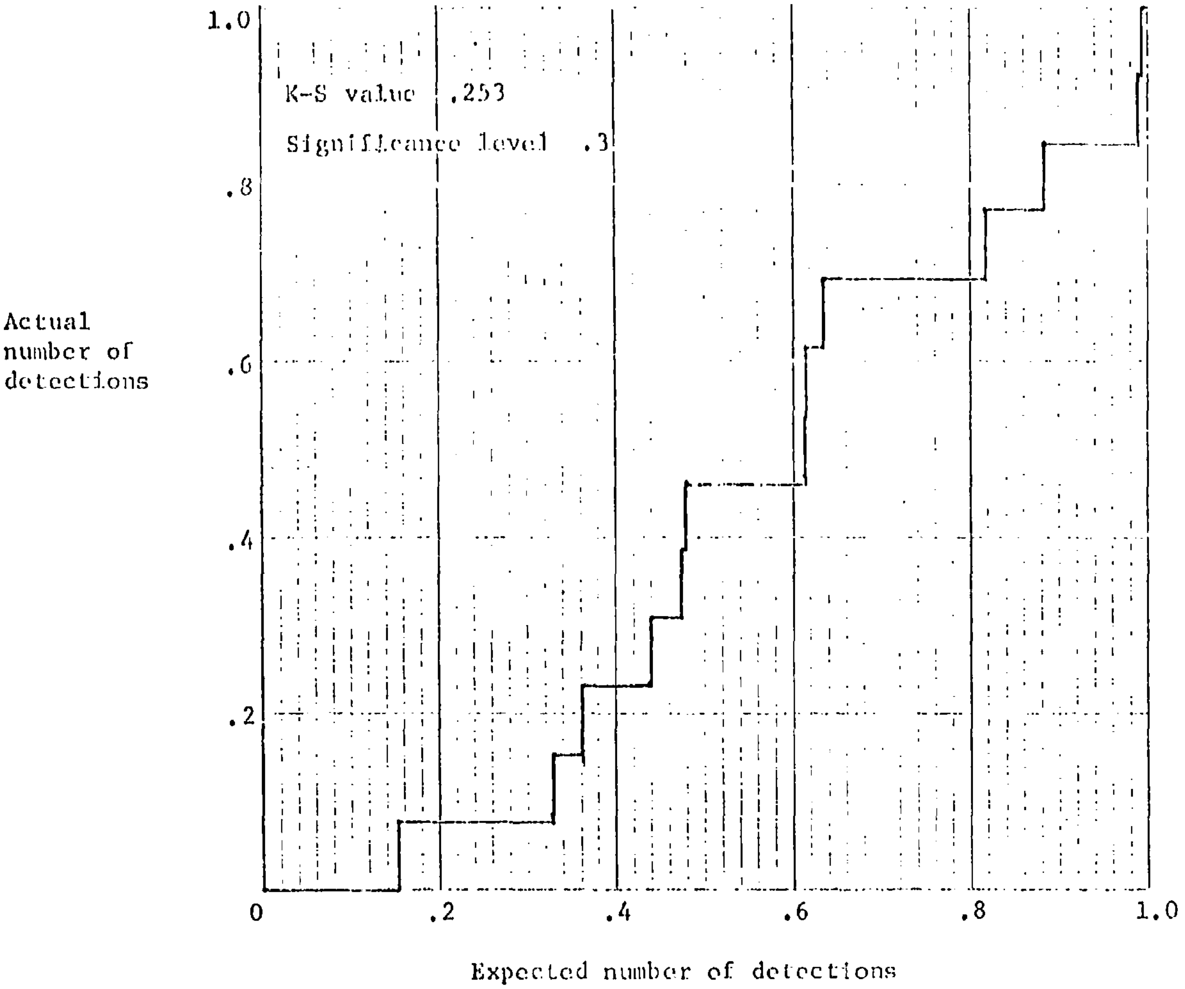


Figure 14. Plot of actual vs. expected number of detections, 30 trials, set A (reversed), range order.

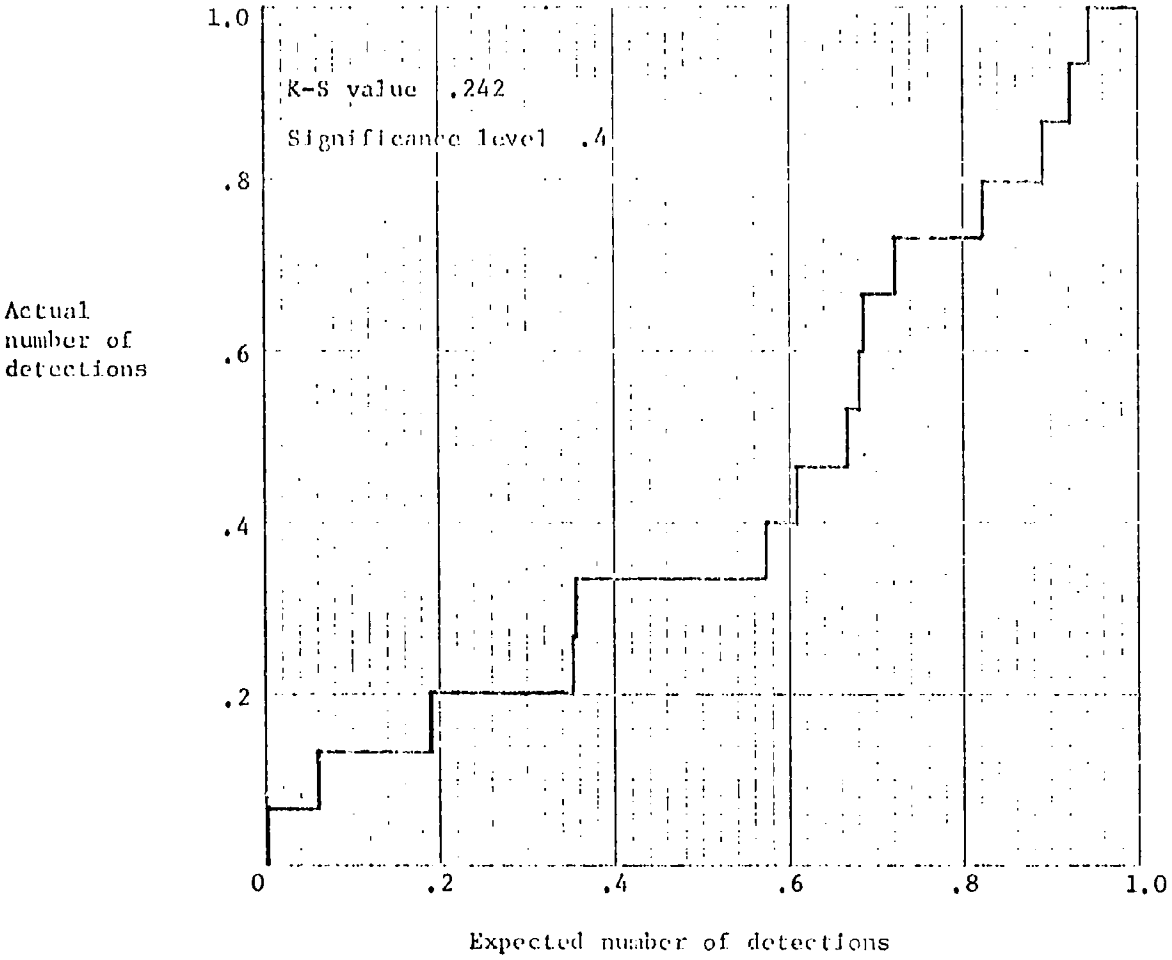


Figure 15. Plot of actual vs. expected number of detections, 30 trials, set B, natural order.

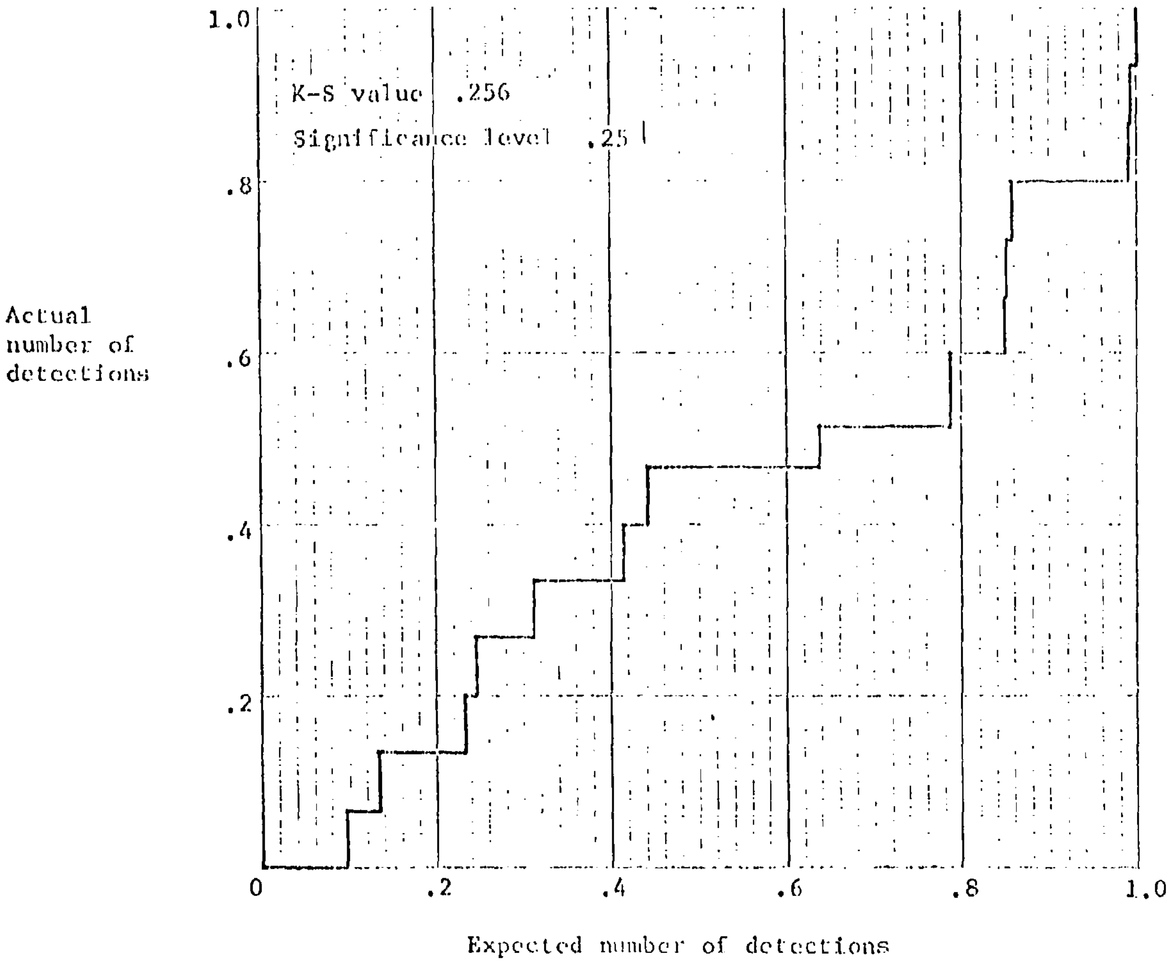


Figure 16. Plot of actual vs. expected number of detections, 30 trials, set B, p_i order.

APPENDIX

Mathematical Notes

A. Combination of two tests of significance. (See Page 10.)

If two independent tests of significance for a given hypothesis have been done, and have resulted in significance levels p_1 and p_2, what significance level has been attained with both together?

ANSWER: Under the null hypothesis, the significance levels attained by two independent tests can be transformed: if p_1 and p_2 are the significance levels, then $z_i = -2 \log_e p_i$ are independent Chi-square variables with two degrees of freedom. Thus $z = z_1+z_2$ is a Chi-square variable with four degrees of freedom, and its significance can be checked with standard Chi-square tables. (It is assumed that the two test statistics are defined so that departure from the null hypothesis in a given direction will push both statistics in the same direction.) See, e.g., Reference [2] for the derivation, which is straightforward.

B. Treatment of points t where $\lambda(t)$ is unbounded. (See Page 15.)

Remember that the immediate desire is to define a function $s(t)$, in terms of which the problem is to be reformulated. Two cases are possible, and are handled differently. First, suppose that the function $\lambda(t)$ is integrable through its (infinite) discontinuity. Then there is no difficulty defining $s(t)$ across the discontinuity, and the problem is solved: $s(t)$ will be continuous, and the remainder of the derivation proceeds unimpeded. Second, suppose the above is not true. Then think of the alternate characterization of $s(t)$, as the expected number of detections by time t. It is, of course, true that $s(t)$ will be discontinuous. The meaning is clear: there is a non-zero probability of a detection exactly at the point t_0 . In this case, the point t_0 is an exceptional point, and can be treated separately: it does not fit the definition of a continuous-time stochastic process, and must be deleted. It can be shown that the number of such points must be, at most, countable, and therefore they can be simply omitted from consideration in terms of the Poisson process. (They can be treated separately, as independent trials: the distribution of the number of detections at such a point must be derived for the hypothesized model, and then the significance of the observed number of detections can be evaluated.)

C. Rearrangement in terms of a suspect variable: proof that the result is still a Poisson process. (See Page 18.)

The intention here is the following: Assuming the truth of the null hypothesis, show that one can break up the trials into segments, rearrange these in order of some other variable r , and still have a realization of a Poisson process.

As shown, for instance, in Reference [10], one can define a non-homogeneous Poisson process consisting of the set of trials joined end-to-end, and then use a new "time" variable s to make it a homogeneous Poisson process with occurrence rate unity. Let S denote the maximum value of the variable s . Define $I_r(s) = 1$ if $r(s) > r$, and $I_r(s) = 0$ otherwise. Let

$$t(r) = \int_0^S I_r(s)\, ds$$

denote the amount of time during which $r(s) > r$. An interval of time (t_1, t_2) in the rearranged set can be considered to be an interval $(t(r_1) , t(r_2))$ as long as t_1 and t_2 are the unique values of t corresponding to some values r_1 and r_2 respectively. (Otherwise, there are trivial complications, which will be considered below.) Then the number of detections in the interval $(t(r_1)\ ,\ t(r_2))$ is precisely the number of detections in $\{s: I_{r_2}(s) - I_{r_1}(s) = 1\}$, which is a Poisson variable, with mean m given by

$$m = \int_0^S (I_{r_2}(s) - I_{r_1}(s))^+ \, ds \ ,$$

where the superscript plus indicates the positive part of the function in parentheses. Finally, since $I_{r_1} = 1$ implies $I_{r_2} = 1$, we can write

$$m = \int_0^S I_{r_2}(s)\, ds - \int_0^S I_{r_1}(s)\, ds$$

$$= t(r_2) - t(r_1)$$

which is the desired result.

The complications that ensue when one or both of the t_i do not correspond uniquely to given r_i are due to the incompleteness of the specification of the ordering. There are many segments which correspond to a given value of r . Once it is specified how these segments are to be ordered among themselves, the same argument used above can be applied to them.

REFERENCES

[1] ABRAMOWITZ, MILTON and STEGUN, IRENE (1964). *Handbook of Mathematical Functions*. (Applied Math. Series No. 55). National Bureau of Standards, Washington, D. C.

[2] ANDERSON, R. L. and BANCROFT, T. A. (1952). *Statistical Theory in Research*. McGraw-Hill Book Co., New York.

[3] ANDERSON, T. W. and DARLING, D. A. (1952). *Ann. Math. Statist*. **23** 193-212.

[4] BARNARD, G. A. (1953) *Biometrika* **40** 212-213

[5] BIRNBAUM, Z. W. (1952). *J. Amer. Statist. Assoc*. **47** 425-441.

[6] BOSSARD, DAVID C. (1970). Statistical tests for detection models. Daniel H. Wagner Associates. AD 702497

[7] LOEVE, MICHEL M. (1963). *Probability Theory*. (3rd ed.). Van Nostrand, Princeton, N. J.

[8] MAGUIRE, B. A., PEARSON E. S. and WYNN, A. H. A. (1953). *Biometrika* **40** 213-216.

[9] OWEN, D. B. (1962). *Handbook of Statistical Tables*. Addison - Wesley, Reading, Massachusetts.

[10] PARZEN, EMANUEL (1962). Stochastic Processes. HolJen-Day, Inc., San Francisco.

[11] PEARSON, KARL (1934). Tables of the Incomplete Γ - Function. Cambridge University Press, Cambridge, Mass.

[12] SIEGEL, SIDNEY (1956). Nonparametric Statistics for the Behavioral Sciences, McGraw-Hill Book Co., New York.